



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 96606

TO: Sheridan Swope  
Location: cm1/12d12/10d01  
Art Unit: 1652  
Friday, June 20, 2003  
Case Serial Number: 966880

From: Paul Schulwitz  
Location: Biotech-Chem Library  
CM1-6B06  
Phone: 305-1954

[paul.schulwitz@uspto.gov](mailto:paul.schulwitz@uspto.gov)

### Search Notes

Examiner Swope,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Paul Schulwitz  
Technical Information Specialist  
STIC Biotech/Chem Library  
(703)305-1954







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 11:07:40 ; Search time 6987 Seconds  
(without alignments)  
12781.153 Million cell updates/sec

Title: US-09-966-880A-9  
Perfect score: 5514  
Sequence: 1 acagacgaatacatgctca.....tcaactcctgactcagag 5514

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 16154066 seqs, 8097743376 residues

Word size : 0

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	272	4.9	293	17 A0042682 CIT-HSP-2
2	87	1.6	693	12 BG757392 602711022
3	84	1.5	541	12 BF238155 601811880
4	84	1.5	743	12 BG686133 602638412
5	84	1.5	942	12 BF975166 602244657
6	71	1.3	953	14 B0065440 AGENCOURT

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
A0042682												
71	1.3	1052	14	BQ055935								
69	1.3	423	14	T06576								
66	1.3	648	17	AG056000								
68	1.2	429	10	AM303243								
68	1.2	578	9	A1811647								
67	1.2	164	12	BF806490								
67	1.2	198	10	BE177917								
67	1.2	256	10	AM562095								
67	1.2	300	9	AA937809								
67	1.2	320	10	BE169901								
67	1.2	331	9	AA582746								
67	1.2	364	9	AA569235								
67	1.2	375	10	AM799144								
67	1.2	415	9	AA224860								
67	1.2	436	10	AM799150								
67	1.2	440	17	A0539921								
67	1.2	444	17	A0056100								
67	1.2	479	9	AA993560								
67	1.2	480	17	A0633793								
67	1.2	503	9	A1376179								
67	1.2	507	10	AM799331								
67	1.2	513	10	BE151195								
67	1.2	532	9	AL601949								
67	1.2	554	10	BE151208								
67	1.2	558	12	BF815902								
67	1.2	566	9	AU147853								
67	1.2	591	10	AM026595								
67	1.2	597	10	AV733437								
67	1.2	613	17	A0389816								
67	1.2	629	12	BG621190								
67	1.2	632	17	A0379336								
67	1.2	651	14	BM993430								
67	1.2	667	17	AG156377								
67	1.2	675	17	A0321493								
67	1.2	677	17	AG158392								
67	1.2	681	9	AL704287								
67	1.2	690	17	AG178815								
67	1.2	692	17	AG173021								
67	1.2	698	17	A0428052								

## ALIGNMENTS

RESULT 1  
A0042682  
LOCUS  
DEFINITION  
CIT-HSP-2326M11.TR CIT-HSP Homo sapiens genomic clone 2326M11, DNA  
sequence.  
ACCESSION  
A0042682  
VERSION  
GI:3310067  
KEYWORDS  
GSS.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 293)  
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Baas,S., Linher,K.,  
Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,  
Simon,M. and Venter,J.C.  
Use of a random BAC End Sequence Database for Sequence-Ready Map  
Building (1998)  
Unpublished (1998)  
Other\_GSSs: CIT-HSP-2326M11.TV  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdamas@tigr.org  
Clones are available from Research Genetics (Info@resgen.com). BAC  
end search page:



```

ORIGIN
Query Match      1.5%; Score 84; DB 12; Length 541;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1035 GAACCATCATTAATTAAGAGATTTTCTGGCCCTGAGACTTCGACGAGGCAAGAAGA 1094
      |||
Db 2 GAACCATCATTAATTAAGAGATTTTCTGGCCCTGAGACTTCGACGAGGCAAGAAGA 61
      |||

Oy 1095 CACTCTGGACACCACTATGACAG 1118
      |||
Db 62 CACTCTGGACACCACTATGACAG 85
      |||

RESULT 4
BG686133      743 bp mRNA linear EST 01-MAY-2001
LOCUS      602638412F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4766234 5',
DEFINITION mRNA sequence.
ACCESSION BG686133
VERSION BG686133.1 GI:13917530
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM1626 row: 9 column: 03
High quality sequence stop: 740.
Location/Qualifiers
1..743
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4766234"
/tissue_type="NIH_MGC_48"
/lab_host="NIH_MGC_48"
/lab_host="primary B-cells from tonsils (cell line)"
/note="Organ: B-cells; Vector: pOR7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT 176 a 197 c 188 g 182 t
ORIGIN

Query Match      1.5%; Score 84; DB 12; Length 743;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1035 GAACCATCATTAATTAAGAGATTTTCTGGCCCTGAGACTTCGACGAGGCAAGAAGA 1094
      |||
Db 2 GAACCATCATTAATTAAGAGATTTTCTGGCCCTGAGACTTCGACGAGGCAAGAAGA 61
      |||

Oy 1095 CACTCTGGACACCACTATGACAG 1118
      |||
Db 62 CACTCTGGACACCACTATGACAG 85
      |||

```

```

RESULT 5
BF975166      942 bp mRNA linear EST 22-JAN-2001
LOCUS      602244657F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4335639 5',
DEFINITION mRNA sequence.
ACCESSION BF975166
VERSION BF975166.1 GI:12342381
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LNCM1207 row: a column: 16
High quality sequence stop: 707.
Location/Qualifiers
1..942
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4335639"
/tissue_type="NIH_MGC_48"
/lab_host="NIH_MGC_48"
/lab_host="primary B-cells from tonsils (cell line)"
/note="Organ: B-cells; Vector: pOR7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT 249 a 209 c 251 g 233 t
ORIGIN

Query Match      1.5%; Score 84; DB 12; Length 942;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 84; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1035 GAACCATCATTAATTAAGAGATTTTCTGGCCCTGAGACTTCGACGAGGCAAGAAGA 1094
      |||
Db 2 GAACCATCATTAATTAAGAGATTTTCTGGCCCTGAGACTTCGACGAGGCAAGAAGA 61
      |||

Oy 1095 CACTCTGGACACCACTATGACAG 1118
      |||
Db 62 CACTCTGGACACCACTATGACAG 85
      |||

RESULT 6
B0065440      953 bp mRNA linear EST 02-APR-2002
LOCUS      AGENCOURT 6855061 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5929977
DEFINITION 5', mRNA sequence.
ACCESSION B0065440
VERSION B0065440.1 GI:19894486
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Mammalia: Eutheria: Primates: Catarrhini: Hominoidea: Homo. 1 (bases 1 to 935)	NIH-MGC	<a href="http://mgc.ncbi.nlm.nih.gov/">http://mgc.ncbi.nlm.nih.gov/</a>	National Institutes of Health, Mammalian Gene Collection (MGC)	
Unpublished (1999)	Contact: Robert Strussberg, Ph.D.	Email: <a href="mailto:cgabbs-remail.nih.gov">cgabbs-remail.nih.gov</a>	Tissue Procurement: Lou Staudt	
CDNA Library Preparation: Rubin Laboratory	DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)	Clone distribution by: Agencourt Bioscience Corporation	found through the I.M.A.G.E. Consortium/LNL at:	
<a href="http://image.llnl.gov">http://image.llnl.gov</a>	Plate: L1CM2108	row: p	column: 10	
High quality sequence stop: 634.	Location/Qualifiers	1..953		

```

FEATURES
    source
        1. 933
            Location/Qualifiers
                sequence_id="1.933"
                sequence_stop="634."
                clone_name="Homo sapiens"
                db_xref="taxon:9606"
                clone_image="5929977"
                clone_lib="NIH_MGC_99"
                clone_type="lymphoma, cell line"
                lab_host="DH10B (phage-resistant)"
                note="Organ: lymph. Vector: pOTR1; Site:1: XhoI; Site:2: EcoRI. CDNA made by oligo-dT priming. Directionally cloned. GGCACGAC(G). Size-selected >500bp for average insert size of 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC library."
        236 c
        233 g
        246 t

```

Query Match	1.38;	Score 71;	DB 14;	Length 933;
Best Local Similarity	100.08;	Pred. No. 0;		
Matches 71;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
Qy	1048 TTGAAGTGAATTTTTCGGCCCTGAGACTTGACGAGGAGCCAAAGACACTCTGGACAC			
Db	1 TTGAAGTGAATTTTTCGGCCCTGAGACTTGACGAGGAGCCAAAGACACTCTGGACAC			
Qy	1108 ACTATGACAC	1118		
Db	61 ACTATGACAC	71		
RESULT 7				
BO055935				
LOCUS				
DEFINITION	BO055935	1052 bp	mrna	linear EST 29-MAR-2007
AGENCOURT	-6796291 NIH-MGC_99	Homo sapiens	CDNA clone	IMAGE:5808181
ACCESSION	BO055935			
VERSION	BO055935.1	GI:19815262		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	1 (bases 1 to 1052)			
TITLE	NIH-MGC <a href="http://mgc.ncl.nih.gov/">http://mgc.ncl.nih.gov/</a>			
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)			
COMMENT	Unpublished (1999)			
	Contact: Robert Strumbe			

```

http://image.llnl.gov
plate: L1CM2051 row: m column: 14
High quality sequence stop: 665.
Location/Qualifiers
1..1052
FEATURES
source
```

FEATURES	Source
ntcpc:///image.llnl.gov Plate: L10CM2051 row: m column: 14 High quality sequence stop: 665. Location/Qualifiers 1..1052 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:5808181" /clone_lib="NIH_MGC_99" /tissue_type="Lymphoma, cell line" /lab_host="DH10B (phage-resistant)" /note="Organ: Lymph; Vector: POTR; Site.1: XhoI; Site.2: into EcoRI/XhoI sites using oligo-dr priming. Directionally cloned GGCAGC(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-CDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."	276 a 250 c 258 g 266 t 2 others

Query Match	1.38;	Score 71;	DB 14;	Length 1052;
Best Local Similarity	100.08;	Pred. No. 0;		
Matches	71;	Conservative 0;	Mismatches 0;	Indels 0;
				Gaps 0;
QY	1048	TTGACGAGCATTTTTCGCGCTGAGACTTCGACGAGGAGGAGAGCACTCTGGACACC		
Db	1	TTGAAATGAGATTTTTCGCGCTGAGACTTCGACGAGGAGGAGAGCACTCTGGACACC	1107	
QY	1108	ACATATGACAGC	1118	
Db	61	ACTATGACAGC	71	

RESULT 8	LOCUS	DEFINITION	EST004665 Fetal brain, Striatum (cat#36206) Homo sapiens cDNA
T06576	T06576	423 bp	EST 30-JUN-1993
ACCESION	T06576	clone HEMDV57 similar to EST containing Alu repeat, mRNA sequence.	
VERSION	T06576.1	GI:317725	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		

REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	COMMENT
1 (bases 1 to 423)	Adams,M.D., Keriavage,A.R., Fields,C. and Venter,J.C.	3,400 expressed sequence tags identify diversity of transcripts from human brain	Nat. Genet. 4, 256-267 (1993)	93364420	Contact: Adams, MD The Institute for Genomic Research 932 Clopper Road, Gaithersburg, MD 20878 Tel: 3018699056 Fax: 3018699423 Email: mtdadams@igf.org Seq primer: M13-21.
					Location/Qualifiers 1. .423

```

FEATURES
SOURCE
    Location/Qualifiers
    1..423
    /organism="Homo sapiens"
    /db_xref="ATCC (inhost):83234"
    /db_xref="taxon:9606"
    /clone="HERDV37"
    /clone_1id="Fetal brain, Stragene (cat#936206)"
    /note="Vector: LambdaZAP-II: 17-18 wk gestation, female
    oligo-dT + random primed cDNA syntheses; lambdaZAP-II
    vector, 1.0kb average inser size."
95 a 109 c 96 g 117 t 6 others
BASE COUNT
ORIGIN

```

Fri Jun 20 09:21:57 2003

us-09-966-880a-9.oli.rst

Query Match 1.3%; Score 69; DB 14; Length 423;  
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;  
 Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

5444 TGTGTGATTTTGTAGATGGGCTTCCACCATGTTGCCAGCGCTGCTCAAACTCC 5503  
 |||||||  
 127 TGTGTGATTTTGTAGATGGGCTTCCACCATGTTGCCAGCGCTGCTCAAACTCC 186  
 |||||||

QY 5444 TGACCTCAG 5512  
 |||||||

DB 187 TGACCTCAG 195

RESULT 9  
 AG056000/c 648 bp DNA linear GSS 02-NOV-2001  
 LOCUS Pan troglodytes DNA, clone: PTB-042C11.R, genomic survey sequence.  
 ACCESSION AG056000  
 VERSION AG056000.1 GI:16593459  
 KEYWORDS Pan troglodytes male lymphoblast DNA, clone: PTB Chimpanzee Male  
 SOURCE BAC library clone: PTB-042C11.R.  
 ORGANISM Pan troglodytes  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pan.

REFERENCE  
 AUTHORS 1 Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
 Totsuki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE BAC and sequences of library PTB  
 JOURNAL Unpublished  
 AUTHORS 2 (bases 1 to 648)  
 Fujiiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,  
 Totsuki, Y., Watanabe, H. and Sakaki, Y.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-NOV-2001) Asao Fujiyama, The Institute of Physical  
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC), Japan  
 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
 (E-mail: chimbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/  
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)  
 Clones are derived from the chimpanzee BAC library PTB This BAC end  
 was generated during the R&D process and may have higher chance of  
 clone tracking errors.

COMMENT  
 Clones are derived from the chimpanzee BAC library PTB This BAC end  
 was generated during the R&D process and may have higher chance of  
 clone tracking errors.

PRIMERS  
 Sequencing: M13Rev

LIBRARY  
 Vector : pKS145  
 R. Site 1 : SacI  
 R. Site 2 : SacI.  
 Location/Qualifiers  
 1. 648  
 /organism="Pan troglodytes"  
 /db\_xref="taxon:9598"  
 /clone="PTB-042C11.R"  
 /sex="male"  
 /cell\_type="lymphoblast"  
 /clone\_id="PTB Chimpanzee Male BAC library"  
 174 a 144 c 177 g 153 t

BASE COUNT  
 174 a 144 c 177 g 153 t

ORIGIN  
 1.3%; Score 69; DB 17; Length 648;  
 Query Match 1.3%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;  
 Best Local Similarity 100.0%; Mismatches 0; Indels 0; Gaps 0;  
 Matches 69; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5439 GCTAATGTTTGTATTTTGTAGATGGGCTTCCACCATGTTGCCAGCGCTGCTCA 5498  
 |||||||  
 DB 558 GCTAATGTTTGTATTTTGTAGATGGGCTTCCACCATGTTGCCAGCGCTGCTCA 499  
 |||||||

QY 5499 ACTGCTGAC 5507  
 |||||||

DB 498 ACTGCTGAC 490

RESULT 10  
 AM303243/c 429 bp mRNA linear EST 18-JAN-2000  
 LOCUS x190C01.x1 NCI CGAP Lu26 Homo sapiens cDNA clone IMAGE:2767392 3'  
 DEFINITION similar to contains Alu repetitive element;; mRNA sequence.  
 ACCESSION AM303243  
 VERSION AM303243.1 GI:6712923  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
 AUTHORS 1 (bases 1 to 429)  
 TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 CONTACT Contact: Robert Strausberg, Ph.D.  
 Email: cgaps-remail.nih.gov  
 CDNA Library Preparation: David B. Krizman, Ph.D.  
 CDNA Library Arrayed by: I.M.A.G.E. Consortium, LNL  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 DNA Distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at:  
 www.bio.lnlnl.gov/dbip/image/image.html  
 Seq primer: -400p from Gibco  
 High quality sequence stop: 424.  
 Location/Qualifiers  
 1. 429  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:2767392"  
 /clone\_id="NCI-CGAP\_Lu26"  
 /tissue\_type="invasive adenocarcinoma"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Organ: lung; Vector: PAMPI; mRNA made from lung  
 adenocarcinoma tissue. cDNA made by oligo-dT priming,  
 directionally cloned. Size-selected on agarose gel,  
 average insert size 500 bp. Primary library,  
 non-amplified."

BASE COUNT  
 138 a 80 c 110 g 101 t

ORIGIN  
 1.2%; Score 68; DB 10; Length 429;  
 Query Match 1.2%; Pred. No. 0; Mismatches 0; Indels 0; Gaps 0;  
 Best Local Similarity 100.0%; Mismatches 0; Indels 0; Gaps 0;  
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTTGTATTTTGTAGATGGGCTTCCACCATGTTGCCAGCGCTGCTCAAACTCTG 5505  
 |||||||  
 DB 349 TTTGTATTTTGTAGATGGGCTTCCACCATGTTGCCAGCGCTGCTCAAACTCTG 290  
 |||||||

QY 5506 ACCTCGA 5513  
 |||||||

DB 289 ACCTCGA 282

RESULT 11  
 A1811647 578 bp mRNA linear EST 15-DEC-1999  
 LOCUS t144c03.x1 NCI CGAP Utl1 Homo sapiens cDNA clone IMAGE:2262532 3'  
 DEFINITION similar to contains Alu repetitive element;; mRNA sequence.  
 ACCESSION A1811647  
 VERSION A1811647.1 GI:5398213  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

REFERENCE  
 AUTHORS 1 (bases 1 to 578)  
 TITLE NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index

JOURNAL  
COMMENT

Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: c9abs-r@emil.nih.gov  
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/RLNL at:  
[www.dio.llnl.gov/dbp/rlnl/image/image.html](http://www.dio.llnl.gov/dbp/rlnl/image/image.html)  
Insert length: 1045 Std Error: 0.00  
Seq primer: -400P from Glibco  
High quality sequence stop: 406.  
Location/Qualifiers

FEATURES  
source

```
1. 578
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Ut1"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/Note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 Kb. Life Technologies catalog #:
11538-014"
```

BASE COUNT 147 a 139 c 114 g 177 t 1 others

## ORIGIN

Query Match 1.28; Score 68; DB 9; Length 578;  
Best Local Similarly 100.0%; Pred. No. 0;  
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5446 TTGTATTTTACAGATGGGTTTCACCATGTTGGCCAGCTGCTCAACCTCTG 5505

Db 130 TTGTATTTTACAGATGGGTTTCACCATGTTGGCCAGCTGCTCAACCTCTG 189

OY 5506 ACCTCAGA 5513

Db 190 ACCTCAGA 197

RESULT 12 164 bp mRNA linear EST 12-JAN-2001  
BF806490/c LOCUS R00-C10037-081100-032-d01\_1 C10037 Homo sapiens cDNA, mRNA

DEFINITION sequence.

ACCESSION BF806490.1 GI:12135479

VERSION EST.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 164)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,

Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-27049922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
([http://www.ludwig.org.br/scripts/gethtml2.pl?rl=RC06ct=RC0-C10037-081100-032-d01\\_1&f3=2000-03-17&t4=1](http://www.ludwig.org.br/scripts/gethtml2.pl?rl=RC06ct=RC0-C10037-081100-032-d01_1&f3=2000-03-17&t4=1))  
Seq primer: puc 18 forward  
High quality sequence stop: 164.  
Location/Qualifiers

FEATURES  
source

```
1. 164
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="C10037"
/dev_stage="Adult"
/Note="Organ: colon; Ins: Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORBESTES PCR (U.S. Letters Patent application
No. 196 716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
```

BASE COUNT 41 a 36 c 52 g 35 t

## ORIGIN

Query Match 1.28; Score 67; DB 12; Length 164;  
Best Local Similarly 100.0%; Pred. No. 0;  
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5446 TTGTATTTTACAGATGGGTTTCACCATGTTGGCCAGCTGCTCAACCTCTG 5505

Db 79 TTGTATTTTACAGATGGGTTTCACCATGTTGGCCAGCTGCTCAACCTCTG 20

OY 5506 ACCTCAG 5512

Db 19 ACCTCAG 13

RESULT 13 198 bp mRNA linear EST 22-JUN-2000  
BE177917/c LOCUS RC3-HT0600-170300-011-g01\_1 HT0600 Homo sapiens cDNA, mRNA

DEFINITION sequence.

ACCESSION BE177917.1 GI:8657069

VERSION EST.

KEYWORDS human.

SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 198)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,

Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-27049922

Fax: +55-11-2707001

Email: [asimpson@ludwig.org.br](mailto:asimpson@ludwig.org.br)

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

([http://www.ludwig.org.br/scripts/gethtml2.pl?rl=RC3-HT0600-170300-011-g01\\_1&f3=2000-03-17&t4=1](http://www.ludwig.org.br/scripts/gethtml2.pl?rl=RC3-HT0600-170300-011-g01_1&f3=2000-03-17&t4=1))

Seq primer: puc 18 forward

High quality sequence stop: 198.

## FEATURES

Location/Qualifiers  
1. 198

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="HT0600"

/dev\_stage="Adult"  
/note="Organ: head, neck; Vector: puc18; Site: 1: Sma1; Site: 2: Sma1; A mini-library was made by cloning products derived from ORESSES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## BASE COUNT

55 a 43 c 59 g 41 t

## ORIGIN

Query Match 1.2%; Score 67; DB 10; Length 198;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTG 5505

Db 79 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTG 20

QY 5506 ACCTCAG 5512

Db 19 ACCTCAG 13

## RESULT 14

AM562095 256 bp mRNA linear EST 09-MAR-2000  
LOCUS AM562095/c

DEFINITION SMOWAFCAP34C01SK Onchocerca volvulus adult female cDNA (SAM98MLM-OVAR) Onchocerca volvulus cDNA clone SMOWAFCAP34C01 5', mRNA sequence.

ACCESSION AM562095  
VERSION AM562095.1 GI:7212211

KEYWORDS EST.  
SOURCE Onchocerca volvulus.  
ORGANISM Onchocerca volvulus.

Eukaryota; Metazoa; Nematoda; Chromadorea; Spirurida; Filarioidea; Onchocercidae; Onchocerca.

REFERENCE 1 (bases 1 to 256)

Li Zotte-Waniewski, M. and Williams, S.A.

Genes expressed in adult female stage of Onchocerca volvulus

Unpublished (1998)

Contact: Steven A. Williams

Molecular Parasitology

Smith College Department of Biological Sciences

Department of Biological Sciences, Clark Science Center, Smith

College, Northampton, MA, 01063, USA

Tel: 4135853826

Fax: 4135853826

Email: genomesmith.edu

Seq primer: pbluescript SK.

Location/Qualifiers

1. 256

/organism="Onchocerca volvulus"

/db\_xref="taxon:6282"

/clone\_lib="SMOWAFCAP34C01"

/clone\_lib="Onchocerca volvulus adult female cDNA

(SAM98MLM-OVAR)"

/sex="female"

/dev\_stage="adult"

/lab\_host="XLI-Blue MRF"

/note="Vector: Lambda uni-zap XR; Site: 1: Eco RI; Site: 2: Xho I; Filarial nematode parasite of humans. Two adult female worms of Onchocerca volvulus were isolated from consenting patients and quick frozen. Adult female mRNA was converted to double-stranded cDNA using reverse transcriptase and oligo(dT) followed by RNase H and DNA

pol I. The library has 7 x 10E5 independent recombinants and the average insert size is ~1100bp. The library was

## BASE COUNT

68 a 60 c 80 g 48 t

## ORIGIN

Query Match 1.2%; Score 67; DB 10; Length 256;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTG 5505

Db 98 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTG 39

QY 5506 ACCTCAG 5512

Db 38 ACCTCAG 32

## RESULT 15

AA937809 300 bp mRNA linear EST 30-APR-1998  
LOCUS AA937809

DEFINITION nW89B02.s1 NCI-CGAP\_P12 Homo sapiens cDNA clone IMAGE:1253735 similar to contains Alu repetitive element; contains element L1 repetitive element ; mRNA sequence.

ACCESSION AA937809.1 GI:3095920

VERSION AA937809

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 300)

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbbs@mail.nih.gov

Tissue Procurement: W. Douglas Figg, Ph.D., Paul H. Duray, M.D.,

Rodriguez F. Chuagui, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: David B. Krizman, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/HLN at:

www.bio.lnlnl.gov/dbip/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham.

Location/Qualifiers

1. 300

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="IMAGE:1253739"

/clone\_lib="NCI-CGAP\_P12"

/sex="male"

/tissue\_type="metastatic prostate bone lesion"

/lab\_host="DH10B"

/note="Vector: PAMP10; mRNA made from metastatic prostate lesion of the bone; cDNA made by oligo-dT priming. Non-directionally cloned. Size-selected on agarose gel, average insert size 600 bp. Library made by D. Krizman, NIH."

BASE COUNT 55 a 84 c 67 g 94 t

## ORIGIN

Query Match 1.2%; Score 67; DB 9; Length 300;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTG 5505

Db 168 TTGTGATTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGCTCAAACTCTG 227

constructed by Michelle Li Zotte-Waniewski with worms provided by Dr. Sara Lustigman. The library is available from Dr. Steven A. Williams, email: genomesmith.edu."

Fri Jun 20 09:21:57 2003

us-09-966-880a-9.oli.rst

Oy	5506	ACCTCAG	5512
Db	228	ACCTCAG	234

Search completed: June 19, 2003, 13:52:28  
Job time : 6992 secs





J. R. Wagner, S. J. Culmann, G. J. Knipfers, R. J. Virol. 65, 3259-3267, 1991  
A: Title: The Q300 gene: a novel transcription unit induced in simian virus 40-infected cells  
A: Reference number: 148725; MUID:91237845; PMID:1851876  
A: Accession: 148725  
A: status: preliminary; translated from GB/EMBL/DBJ  
A: molecule type: DNA  
A: Residues: 1-77 <RES>  
A: Cross-references: EMBL:X52164; NID:953860; PIDN:CA936417.1; PID:953861

Alignment Scores:	
Pred. No.:	3.9e-05
Score:	14.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	0.81%
DB:	2
Length:	77
Matches:	77
Conservative:	14
Mismatches:	0
Indels:	0
Gaps:	0

US-09-966-880A-9 (1-5514) x I48725 (1-77)

[illegible]

RESULT 3  
S01441  
class II histocompatibility antigen HLA-DR beta-1 chain precursor (clone LS1.1.1) - huma

C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 21-Jan-2000  
C:Accession: S01441  
R:Caltrns, J.S.; Dahl, C.A.; Curtisinger, J.M.; Bach, F.H.  
Molecular Acids Pos 15 0252 1000

A: Molecule type: mRNA  
A: Title: Identification of a novel DR beta cDNA clone.  
A: Reference number: S01441; MID: 89016649; PMID: 31744622  
A: Accession: S01441

A:Residues: 1296 <CAI>  
A:Cross-references: EMBL:X12544; NID:g32210; PIDN:CAA31061.1; PID:g32211  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1968  
A:Note: only part of the sequence is shown

A:Note: the authors translated the codon TTT for residue 266 as Gly  
C:Superfamily: class II histocompatibility antigen: immunoglobulin homology  
F:1-29/Domain: signal sequence #status predicted <IG>  
E:30-296/Product: class II histocompatibility antigen, HLA-DR beta-1 chain  
F:139-204/Domain: immunoglobulin homology <IMM>

Alignment Scores:	
Pred. No.:	0.000426
Score:	13.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	0.75%
DB:	2
Length:	296
Matches:	13
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-966-880A-9 (1-5514) x S01441 (1-296)

QY	3568	TTGGCTCACGCGCTGTATCCACGACTCTGGGAGGCTGAG	3606
Db	280	TriPleuthrProValIleProIaleuTripcIuaIaGlu	292

RESULT 4  
E40301

C:Species: Homo sapiens (man) - human

C;Accession: F40201

personal communication, 1992

A;Reference number: A40201

A;MOLECULE TYPE: DN  
A;ACCESSION: F40201

A:ResIdues: 1-673 &lt;CLA&gt;

R;Claverie, J.M.

Genomics 12, 838-841, 1992  
 ATitle: Identifying coding exons by similarity search: Alu-derived and other potential  
 A:Reference number: A40200; MUID: 92241891; PMID:1572661  
 A:Contents: annotation  
 C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of  
 in-frame stop codons as shown as 'x'.  
 C:Comment: Any significant similarity of a predicted protein sequence to a portion of

Alignment Scores:	
Pred. No.:	0.000409
Score:	13.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	0.75%
DB:	4
Length:	673
Matches:	13
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

US-09-966-880A-9 (1-5514) x F40201 (1-673)

```
QY      3562 GTGTGGTGGCTACGCGCTGTATCCAGCACTCTGGAG 3600
          |||
Db       3 ValTrpTrpLeuthrProValIleProAlaLeuTrpGlu 15
```

RESULT 5  
T18321  
hypothetical protein L7610.12 - Leishmania major  
C:Snodgrass, Leishmania major

CDate: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 11-Jan-2002  
CAccession: T18321  
R.Oliver, K.; Murphy, L.; Quail, M.; Lawson, D.; Harris, D.; Rajandream, M.  
Submitted to the EMBL Data Library, Nov. 1999

A:Reference number: Z18876  
A:Accession: T18321  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Accession: T18321

A;Residues: 1-169 <OLI>  
A;Cross-references: EMBL:AL034356; NID:e1371878; PIDN:CAA22247.1  
C;Genetics:  
A;Mol. Cr. 1761013

Alignment Scores  
Pred. No.:  
Score:

Pred. No.:	0.005	Length:	159
Score:	12.00	Matches:	12
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.69%	Indels:	0
DB:	2	Gaps:	0

US-09-966-880A-9 (1-5514) x T18321 (1-169)

```

Oy      2840 ACACACACACACACACACACACACACACAC 2877
          |||||
Db      120 ThriThriThriThriThriThriThri 131

```

RESULT  
S57138

N;Alternate names: hypothetical protein J2027  
hypothetical protein J0K15W - yeast (Saccharomyces cerevisiae),  
hypothetical protein J0K15W - yeast (Saccharomyces cerevisiae),

```
C>Date: 23-Aug-1995 #sequence_revision 08-Sep-1995 #text_change 19-Apr-2002
```

R;Rose, M.; Koetter, P.; Entian, K.D.  
submitted to the Protein Sequence Database  
October 1995

A: Accession: S57138

A; molecule type: DNA

A/Residues: 1-169 <ROS>  
A/Cross-references: EMBL:Z40615, NTD:c1015823, PIDN:CA190645.1, PTD:c1015823, CSDDB:C

C;Genetics:

A;Gene: MIPS:YJR115w

```
A;Cross-references:
A:Map position: 10R
```

10/10/1944

## Alignment Scores:

Pred. No.: 0.005 Length: 169  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.69% Indels: 0  
DB: 2 Gaps: 0

US-09-966-880A-9 (1-5514) x S57138 (1-169)

QY 2840 ACACACACACACACACACACACACACACACAC 2875

Db 18 ThriHsthrHsthrHsthrHsthrHsthrHsthrHst 29

## RESULT 7

B40201 artifact-warning sequence (translated ALU class B) - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000

C:Accession: B40201

R:Claverie, J.M.

Personal communication, 1992

A:Reference number: A40201

A:Accession: B40201

A:Molecule type: DNA

A:Residues: 1-301 &lt;CLAA&gt;

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potential

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c

in-frame stop codons are shown as 'x'.  
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

## Alignment Scores:

Pred. No.: 0.00486 Length: 301  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.69% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x B40201 (1-301)

QY 3614 CCACCCACCTCAGCCTCCAGAGTGTGGATTACA 3579

Db 283 ProProthrsrAlasrGlnserAlaglylththr 294

## RESULT 8

C40201 artifact-warning sequence (translated ALU class C) - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000

C:Accession: C40201

R:Claverie, J.M.

Personal communication, 1992

A:Reference number: A40201

A:Accession: C40201

A:Molecule type: DNA

A:Residues: 1-613 &lt;CLAA&gt;

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potential

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c

in-frame stop codons are shown as 'x'.  
C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

## Alignment Scores:

Pred. No.: 0.0047 Length: 613  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.69% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x C40201 (1-613)

QY 3568 TGCTCAGCCTGTATCCACGACACTGTGGAGCT 3603

Db 5 TrrpleuthrProValleProAlaleuthrplua1a 16

## RESULT 9

C40201 artifact-warning sequence (translated ALU class C) - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000

C:Accession: C40201

R:Claverie, J.M.

Personal communication, 1992

A:Reference number: A40201

A:Accession: C40201

A:Molecule type: DNA

A:Residues: 1-613 &lt;CLAA&gt;

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potentl

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames o

in-frame stop codons are shown as 'x'.  
C:Comment: Any significant similarity of a predicted protein sequence to a portion of

## Alignment Scores:

Pred. No.: 0.0047 Length: 613  
Score: 12.00 Matches: 12  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.69% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x C40201 (1-613)

QY 3602 GCCTCCAGAGTGTGGATTACAGCGCTGAGCCAC 3567

Db 599 AlasrGlnserAlaglylththrGlyValserHs 610

## RESULT 10

T42696 hypothetical protein DKF2p434G1035.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-Jan-2000

C:Accession: T42696

R:Postula, A.; Klein, M.; Mewes, H.W.; Gaassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, November 1999

A:Reference number: 222232

A:Accession: T42696

A:Molecule type: mRNA

A:Residues: 1-163 &lt;AAA&gt;

A:Cross-references: EMBL:AL133039

A:Experimental source: adult testis; clone DKF2p434G1035

C:Genetics:

A:Note: DKF2p434G1035.1

## Alignment Scores:

Pred. No.: 0.0573 Length: 163  
Score: 11.00 Matches: 11  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.63% Indels: 0



Db 442 LeuLysLysIleArgArgLysIleLysAsnLys 452

## RESULT 15

E40201

artifact-warning sequence (translated ALU class E) - human

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000

C:Accession: E40201

R:Claverie, J.M.

personal communication, 1992

A:Reference number: A40201

A:Accession: E40201

A:Molecule type: DNA

A:Residues: 1-597 &lt;CLAV&gt;

R:Claverie, J.M.

Genomics 12, 838-841, 1992

A:Title: Identifying coding exons by similarity search: Alu-derived and other potential

A:Reference number: A40200; MUID:92241891; PMID:1572661

A:Contents: annotation

C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c

in-frame stop codons are shown as 'X'.

C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

## Alignment Scores:

Pred. No.:	0.0537	Length:	597
Score:	11.00	Matches:	11
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	0.63%	Indels:	0
DB:	4	Gaps:	0

US-09-966-880A-9 (1-5514) x E40201 (1-597)

QY 3558 CCAGGTGTGGTGCACGCTGTAATCCAGC 3590

Db 201 ProGlyValValAlaHisAlaCysAsnProSer 211

Search completed: June 14, 2003, 19:35:41  
 Job time : 181.5 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 07:46:55 ; Search time 216 Seconds  
(without alignments)  
7828.776 Million cell updates/sec

Title: US-09-966-880A-9  
Perfect score: 5514  
Sequence: 1 acagacgaatacatgtgtcca.....tcaaacctctgacctcagag 5514

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 441362 seqs, 153338381 residues

Word size : 0

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67	1.2	20303	1 US-08-370-975B-6	Sequence 6, Appli
2	67	1.2	26764	1 US-08-370-975B-1	Sequence 1, Appli
3	66	1.2	8758	4 US-09-799-345-3	Sequence 3, Appli
4	63	1.1	99500	4 US-09-798-096-10	Sequence 10, Appl
5	61	1.1	2559	2 US-08-724-774B-3	Sequence 3, Appli
6	61	1.1	2559	3 US-09-089-595-3	Sequence 3, Appli
7	61	1.1	2559	4 US-09-382-855-3	Sequence 3, Appli
8	61	1.1	2559	4 US-09-183-714B-3	Sequence 3, Appli
9	61	1.1	2559	4 US-09-642-281-3	Sequence 3, Appli
10	61	1.1	43950	4 US-09-735-934A-3	Sequence 3, Appli
11	60	1.1	1541	6 5183884-1	Patent No. 5183884
12	60	1.1	1542	1 US-07-978-895-1	Sequence 1, Appli
13	60	1.1	1542	1 US-08-473-119-1	Sequence 1, Appli
14	60	1.1	1542	2 US-08-475-352-1	Sequence 3, Appli
15	58	1.1	36159	4 US-09-749-588-3	Sequence 382, App
16	56	1.0	57	1 US-08-222-177A-382	Sequence 49, Appl
17	56	1.0	4576	2 US-08-832-877-49	Sequence 49, Appl
18	55	1.0	4576	2 US-08-832-877-49	Sequence 49, Appl
19	55	1.0	36741	4 US-09-301-665-3	Sequence 3, Appli
20	54	1.0	36651	4 US-09-738-894A-3	Sequence 3, Appli
21	54	1.0	48974	4 US-08-920-422-17	Sequence 17, Appl
22	54	1.0	112132	4 US-09-741-150-3	Sequence 3, Appli
23	52	0.9	3474	4 US-09-527-236A-1	Sequence 1, Appli
24	52	0.9	6140	4 US-09-605-785-536	Sequence 536, App
25	52	0.9	6140	4 US-09-439-313-536	Sequence 536, App
26	51	0.9	29629	4 US-09-729-995-3	Sequence 3, Appli
27	51	0.9	59065	4 US-09-813-817-3	Sequence 3, Appli

c 28	51	0.9	59065	4 US-09-978-197-3	Sequence 3, Appli
c 29	51	0.9	81001	4 US-09-750-580-1	Sequence 1, Appli
c 30	51	0.9	98844	4 US-09-791-211-10	Sequence 10, Appl
c 31	51	0.9	152331	3 US-09-128-155-16	Sequence 16, Appl
c 32	51	0.9	162450	4 US-09-345-882-1	Sequence 1, Appli
c 33	51	0.9	162450	4 US-09-345-882-1	Sequence 1, Appli
c 34	51	0.9	176373	3 US-09-128-155-17	Sequence 17, Appl
c 35	50	0.9	807	2 US-08-531-927B-9	Sequence 9, Appli
c 36	50	0.9	1776	2 US-08-531-927B-1	Sequence 12, Appl
c 37	50	0.9	1776	4 US-09-041-886-12	Sequence 12, Appl
c 38	50	0.9	6769	1 US-08-480-784-20	Sequence 20, Appl
c 39	50	0.9	6769	1 US-08-483-553-20	Sequence 20, Appl
c 40	50	0.9	6769	1 US-08-487-002-20	Sequence 20, Appl
c 41	50	0.9	6769	1 US-08-483-554B-20	Sequence 20, Appl
c 42	50	0.9	6769	1 US-08-488-011B-20	Sequence 20, Appl
c 43	50	0.9	6769	4 US-08-850-727-20	Sequence 20, Appl
c 44	50	0.9	6769	5 PCT-US95-10202-20	Sequence 20, Appl
c 45	50	0.9	6769	5 PCT-US95-10203-20	Sequence 20, Appl

ALIGNMENTS

RESULT 1  
US-08-370-975B-6  
; Sequence 6, Application US/08370975B  
; Patent No. 5622851  
; GENERAL INFORMATION:  
; APPLICANT: Maley, Frank  
; APPLICANT: Maley, Gladys F.  
; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene  
; NUMBER OF SEQUENCES: 14  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
; STREET: Clinton Square, P.O. Box 1051  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 14603  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/370,975B  
; FILING DATE: 10-JAN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Timilan, Susan J.  
; REGISTRATION NUMBER: 34,103  
; REFERENCE/DOCKET NUMBER: 20894/80  
; TELEPHONE: (716)263-1636  
; TELEFAX: (716)263-1600  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 20303 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; POSITION IN GENOME:  
; CHROMOSOME/SEGMENT: 4q35

Query Match 1.2%; Score 67; DB 1; Length 20303;  
Best Local Similarity 100.0%; Pred. No. 2.6e-16;  
Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5446 TTTGTATTTTAGTAGAGATGGGTTTCACCATGTTGCCAGGCTGCTCAAACTCTG 5505  
|||||

Db 6392 TTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 6451  
 QY 5506 ACCTCAG 5512  
 Db 6452 ACCTCAG 6458

RESULT 2

US-08-370-975B-1  
 ; Sequence 1, Application US/08370975B  
 ; Patent No. 5622851  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Maley, Frank  
 ; APPLICANT: Maley, Gladys F.  
 ; APPLICANT: Weiner, Karen X.B.  
 ; TITLE OF INVENTION: Human Deoxycytidylate Deaminase Gene  
 ; NUMBER OF SEQUENCES: 14  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Nixon, Hargrave, Devans & Doyle  
 ; STREET: Clinton Square, P.O. Box 1051  
 ; CITY: Rochester  
 ; STATE: New York  
 ; COUNTRY: USA  
 ; ZIP: 14603

COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patent In Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/370,975B  
 ; FILING DATE: 10-JAN-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Timian, Susan J.  
 ; REGISTRATION NUMBER: 34,103  
 ; REFERENCE/DOCKET NUMBER: 20894/80  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (716)263-1636  
 ; TELEFAX: (716)263-1600  
 ; INFORMATION FOR SEQ ID NO: 1:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 26764 base pairs  
 ; TYPE: nucleic acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: DNA (genomic)  
 ; POSITION IN GENOME:  
 ; CHROMOSOME/SEGMENT: 4q35  
 ; US-08-370-975B-1

Query Match 1.2%; Score 67; DB 1; Length 26764;  
 Best Local Similarity 100.0%; Pred. No. 2.5e-16;  
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5446 TTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 5505  
 Db 8355 TTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 8414  
 QY 5506 ACCTCAG 5512  
 Db 8415 ACCTCAG 8421

RESULT 3

US-09-799-345-3  
 ; Sequence 3, Application US/09799345  
 ; Patent No. 6323016  
 ; GENERAL INFORMATION:  
 ; APPLICANT: YE, Jane et al.  
 ; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
 ; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
 ; TITLE OF INVENTION: THEREOF

FILE REFERENCE: CL001156  
 ; CURRENT APPLICATION NUMBER: US/09/799,345  
 ; CURRENT FILING DATE: 2001-03-06  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 3  
 ; LENGTH: 8758  
 ; TYPE: DNA  
 ; ORGANISM: Human  
 ; US-09-799-345-3

Query Match 1.2%; Score 66; DB 4; Length 8758;  
 Best Local Similarity 100.0%; Pred. No. 6.9e-16;  
 Matches 66; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5446 TTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 5505  
 Db 6196 TTGTATTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 6255  
 QY 5506 ACCTCA 5511  
 Db 6256 ACCTCA 6261

RESULT 4

US-09-798-096-10/c  
 ; Sequence 10, Application US/09798096  
 ; Patent No. 6399378  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Donna T. Ward  
 ; APPLICANT: Andrew T. Watt  
 ; TITLE OF INVENTION: ANTISENSE MODULATION OF RECQL2 EXPRESSION  
 ; FILE REFERENCE: RTS-0207  
 ; CURRENT APPLICATION NUMBER: US/09/798,096  
 ; CURRENT FILING DATE: 2001-03-01  
 ; NUMBER OF SEQ ID NOS: 89  
 ; SEQ ID NO 10  
 ; LENGTH: 99500  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; US-09-798-096-10

Query Match 1.1%; Score 63; DB 4; Length 99500;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-15;  
 Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 5450 TATTTTATAGATAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 5509  
 Db 65234 TATTTTATAGATAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG 65175  
 QY 5510 CAG 5512  
 Db 65174 CAG 65172

RESULT 5

US-08-724-774B-3  
 ; Sequence 3, Application US/08724774B  
 ; Patent No. 5908778  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coullie,  
 ; APPLICANT: Pierre; Cerrozzini, Jean-Charles; Carrel,  
 ; APPLICANT: Stefan; Reed, Daryl  
 ; TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor  
 ; TITLE OF INVENTION: Rejection Antigen Precursors Mage-10,  
 ; TITLE OF INVENTION: Antibodies Specific To The Molecule, and  
 ; TITLE OF INVENTION: Uses Thereof  
 ; NUMBER OF SEQUENCES: 5  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Felfe & Lynch  
 ; STREET: 805 Third Avenue  
 ; CITY: New York City



STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/724,774B  
FILING DATE: 03-October-1996  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Hanson, No. 5908778man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 5457  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2559 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-724-774B-3

Query Match 1.1% Score 61; DB 2; Length 2559;  
Best Local Similarity 100.0%; Pred. No. 6.1e-14;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5452 TTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCA 5511  
|||||  
Db 2202 TTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCA 2261  
Qy 5512 G 5512  
|  
Db 2262 G 2262

RESULT 6  
US-09-089-595-3  
Sequence 3, Application US/09089595  
Patent No. 6153728  
GENERAL INFORMATION:  
APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coulie, Pierre;  
APPLICANT: Pierre; Cerrottini, Jean-Charles; Carrel,  
APPLICANT: Stefan; Reed, Daryl  
TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor  
TITLE OF INVENTION: Rejection Antigen Precuros Mage-10,  
TITLE OF INVENTION: Antibodies Specific To The Molecule, and  
TITLE OF INVENTION: Uses Thereof  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Felfe & Lynch  
STREET: 805 Third Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10022  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/089,595  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/724,774  
FILING DATE: 03-October-1996  
ATTORNEY/AGENT INFORMATION:

NAME: Hanson, No. 6153728man D.  
REGISTRATION NUMBER: 30,946  
REFERENCE/DOCKET NUMBER: LUD 5457  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2559 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-089-595-3

Query Match 1.1% Score 61; DB 3; Length 2559;  
Best Local Similarity 100.0%; Pred. No. 6.1e-14;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5452 TTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCA 5511  
|||||  
Db 2202 TTTTGTAGAGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCA 2261  
Qy 5512 G 5512  
|  
Db 2262 G 2262

RESULT 7  
US-09-382-855-3  
Sequence 3, Application US/09382855  
Patent No. 6174692  
GENERAL INFORMATION:

APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coulie, Pierre;  
APPLICANT: Cerrottini, Jean-Charles; Carrel, Stefan; Reed, Daryl  
TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor Rejection  
TITLE OF INVENTION: Antigen Precuros Mage-10, Antibodies Specific To The Molecu  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski LLP  
STREET: 666 Fifth Avenue  
CITY: New York City  
STATE: New York  
COUNTRY: USA  
ZIP: 10103

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage  
COMPUTER: IBM  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/382,855  
FILING DATE: 25-August-1999  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/089,595  
FILING DATE: 02-June-1998  
APPLICATION DATA:  
APPLICATION NUMBER: US 08/724,774  
FILING DATE: 03-October-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Schofield, Mary Anne  
REGISTRATION NUMBER: 36,669  
REFERENCE/DOCKET NUMBER: LUD 5457.2 DIV - JEL/MAS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 318-3000  
TELEFAX: (212) 752-5958  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2559 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-09-382-855-3

Query Match 1.1%: Score 61; DB 4; Length 2559;  
Best Local Similarity 100.0%; Pred. No. 6.1e-14;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5452 TTTTGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCA 5511  
|||||  
Db 2202 TTTTGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCA 2261

QY 5512 G 5512

Db 2262 G 2262

## RESULT 8

US-09-183-714B-3  
; Sequence 3, Application US/09183714B  
; Patent No. 6221593  
; GENERAL INFORMATION:  
; APPLICANT: Boon-Falleur, Thierry  
; APPLICANT: Brasseur, Francis  
; APPLICANT: Rimoldi, Donata  
; APPLICANT: De Plaen, Etienne  
; TITLE OF INVENTION: Method for Determining Cancer by Determining Expression  
; TITLE OF INVENTION: Of MAGE-10  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/09/183,714B  
; PRIOR FILING DATE: 1998-10-30  
; PRIOR APPLICATION NUMBER: US 08/724,774  
; PRIOR FILING DATE: 1998-10-03  
; NUMBER OF SEQ ID NOS: 7  
; SEQ ID NO 3  
; LENGTH: 2559  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
US-09-183-714B-3

Query Match 1.1%: Score 61; DB 4; Length 2559;  
Best Local Similarity 100.0%; Pred. No. 6.1e-14;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5452 TTTTGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCA 5511  
|||||  
Db 2202 TTTTGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCA 2261

QY 5512 G 5512

Db 2262 G 2262

## RESULT 9

US-09-642-281-3  
; Sequence 3, Application US/09642281  
; Patent No. 6387698  
; GENERAL INFORMATION:  
; APPLICANT: Rimoldi, Donata; Jongeneel, Victor; Coullie, Pierre;  
; Cerrotoni, Jean-Charles; Carrel, Stefan; Reed, Daryl  
; TITLE OF INVENTION: MAGE-10 ENCODING cDNA, The Tumor Rejection  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Fulbright & Jaworski LLP  
; STREET: 666 Fifth Avenue  
; CITY: New York City  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.5 inch, 144 kb storage  
; COMPUTER: IBM  
; OPERATING SYSTEM: PC-DOS  
; SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/642,281  
FILING DATE: 18-Aug-2000  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/089,595  
FILING DATE: 02-June-1998  
APPLICATION NUMBER: US 08/724,774  
FILING DATE: 03-October-1996

ATTORNEY/AGENT INFORMATION:  
NAME: Schofield, Mary Anne  
REGISTRATION NUMBER: 36,869  
REFERENCE/DOCKET NUMBER: LUD 5457.2 DIV - JEL/MAS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 318-3000  
TELEFAX: (212) 752-5958

## INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2559 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 3:

US-09-642-281-3

Query Match 1.1%: Score 61; DB 4; Length 2559;  
Best Local Similarity 100.0%; Pred. No. 6.1e-14;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5452 TTTTGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCA 5511  
|||||  
Db 2202 TTTTGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCTCA 2261

QY 5512 G 5512

Db 2262 G 2262

## RESULT 10

US-09-735-934A-3/C  
; Sequence 3, Application US/09735934A  
; Patent No. 6372468  
; GENERAL INFORMATION:  
; APPLICANT: Li, Jiayin et al  
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC  
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES  
; TITLE OF INVENTION: THEREOF  
; FILE REFERENCE: CLO00851  
; CURRENT APPLICATION NUMBER: US/09/735,934A  
; CURRENT FILING DATE: 2000-12-14  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 43950  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-735-934A-3

## Query Match

Best Local Similarity 100.0%; Pred. No. 3.8e-14;  
Matches 61; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5450 TATTTTGTAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCT 5509  
|||||

Db 35406 TATTTTGTAGATGGGTTTCCACCATGTTGCCAGGCTGGTCTCAAACTCCTGACCT 35347

QY 5510 C 5510

Db 35346 C 35346

## RESULT 11

5183884-1

```
; Patent No. 5183884
; APPLICANT: KRAUS, MATTHIAS H.; AARONSON, STUART A.
; TITLE OF INVENTION: DNA SEGMENT ENCODING A GENE FOR A
; RECEPTOR RELATED TO THE EPIDERMAL GROWTH FACTOR RECEPTOR
; NUMBER OF SEQUENCES: 5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/444,406
; FILING DATE: 01-DEC-1989
; SEQ ID NO:1:
; LENGTH: 1541
5183884-1

Query Match      1.1%; Score 60; DB 6; Length 1541;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5453 TTTTAGTAGAGATGGGTTTCCACCATGTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 5512
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1410 TTTTAGTAGAGATGGGTTTCCACCATGTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 1469

RESULT 12
US-07-978-895-1
; Sequence 1, Application US/07978895
; Patent No. 5480968
; GENERAL INFORMATION:
; APPLICANT: Kraus, Matthias H.
; APPLICANT: Aaronson, Stuart A.
; TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE
; TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREO, AND
; TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THEREO
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 400
; STREET: 133 Carnegie Way, N.W.
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.A.
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/978,895
; FILING DATE: 19921110
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/444,406
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Perryman, David G.
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1542 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 66..221
; FEATURE:
; NAME/KEY: exon
; LOCATION: 780..855
; NAME/KEY: exon
```

```
; LOCATION: 1040..1185
; FEATURE:
; NAME/KEY: intion
; LOCATION: 222..779
; FEATURE:
; NAME/KEY: intron
; LOCATION: 856..1039
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(66..221, 780..855, 1040..1185)
US-07-978-895-1

Query Match      1.1%; Score 60; DB 1; Length 1542;
Best Local Similarity 100.0%; Pred. No. 1.5e-13;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5453 TTTTAGTAGAGATGGGTTTCCACCATGTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 5512
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1411 TTTTAGTAGAGATGGGTTTCCACCATGTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 1470

RESULT 13
US-08-473-119-1
; Sequence 1, Application US/08473119
; Patent No. 5820859
; GENERAL INFORMATION:
; APPLICANT: Kraus, Matthias H.
; APPLICANT: Aaronson, Stuart A.
; TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE
; TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREO, AND
; TITLE OF INVENTION: BIOASSAYS AND METHODS RELATED THEREO
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 400
; STREET: 133 Carnegie Way, N.W.
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.A.
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/473,119
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/978,895
; FILING DATE: 10-NOV-1992
; APPLICATION NUMBER: US 07/444,406
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Perryman, David G.
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 66..221
; FEATURE:
; NAME/KEY: exon
; LOCATION: 780..855
```

```

; FEATURE:
; NAME/KEY: exon
; LOCATION: 1040..1185
; FEATURE:
; NAME/KEY: intron
; LOCATION: 222..779
; FEATURE:
; NAME/KEY: intron
; LOCATION: 856..1039
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(66..221, 780..855, 1040..1185)
US-08-473-119-1

Query Match
Best Local Similarity 1.1%; Score 60; DB 1; Length 1542;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5453 TTTTAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 5512
Db 1411 TTTTAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 1470

RESULT 14
US-08-475-352-1
; Sequence 1, Application US/08475352
; Patent No. 5916755
; GENERAL INFORMATION:
; APPLICANT: Kraus, Matthias H.
; APPLICANT: Aaronson, Stuart A.
; TITLE OF INVENTION: AN ISOLATED POLYPEPTIDE RELATED TO THE
; TITLE OF INVENTION: EPIDERMAL GROWTH FACTOR RECEPTOR, ANTIGEN THEREO, AND
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Suite 400
; STREET: 133 Carnegie Way, N.W.
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: U.S.A.
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475.352
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/978,895
; FILING DATE:
; APPLICATION NUMBER: US 07/444,406
; FILING DATE: 01-DEC-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Perryman, David G.
; REGISTRATION NUMBER: 33,438
; REFERENCE/DOCKET NUMBER: 1414-028
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1542 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: exon
; LOCATION: 66..221
; FEATURE:

```

```

; NAME/KEY: exon
; LOCATION: 780..855
; FEATURE:
; NAME/KEY: exon
; LOCATION: 1040..1185
; FEATURE:
; NAME/KEY: intron
; LOCATION: 222..779
; FEATURE:
; NAME/KEY: intron
; LOCATION: 856..1039
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(66..221, 780..855, 1040..1185)
US-08-475-352-1

Query Match
Best Local Similarity 1.1%; Score 60; DB 2; Length 1542;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5453 TTTTAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 5512
Db 1411 TTTTAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 1470

RESULT 15
US-09-749-588-3
; Sequence 3, Application US/09749588
; Patent No. 6423521
; GENERAL INFORMATION:
; APPLICANT: CHANDRAMOULISWARAN, Ishwar et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: THEREOF
; FILE REFERENCE: CL001068
; CURRENT APPLICATION NUMBER: US/09/749,588
; CURRENT FILING DATE: 2000-12-28
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 36159
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)..(36159)
; OTHER INFORMATION: n = A,T,C or G
US-09-749-588-3

Query Match
Best Local Similarity 1.1%; Score 58; DB 4; Length 36159;
Matches 58; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5450 TATTTTAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGAC 5507
Db 11646 TATTTTAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGAC 11703

Search completed: June 19, 2003, 11:55:48
Job time : 219 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 14, 2003, 18:48:04 ; Search time 69.5 Seconds

(without alignments)  
6581.309 Million cell updates/sec

Title: US-09-966-880A-9

Perfect score: 1736

Sequence: 1 acagacgaatacatagtgtcca.....tcaaactcctgacctcagag 5514

Scoring table:

OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Word size: 1

Total number of hits satisfying chosen parameters: 225614

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=xlp  
-O=/cgn2\_1/USPTO\_spool/US09966880/runat\_14062003\_175620\_11056/app\_query.fasta\_1.5703  
-DB=swissprot\_40 -QFMT=fastan -SUFFIX=olin2p.rsp -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09966880.scgn.1\_118\_erunat\_14062003\_175620\_11056 -NCPU=3  
-NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	0.9	585	1 ALU5_HUMAN	P39192 homo sapien
2	16	0.9	591	1 ALU8_HUMAN	P39195 homo sapien
c 3	16	0.9	591	1 ALU8_HUMAN	P39195 homo sapien
4	16	0.9	593	1 ALU7_HUMAN	P39194 homo sapien
c 5	16	0.9	593	1 ALU7_HUMAN	P39194 homo sapien
c 6	14	0.8	77	1 Q300_MOUSE	Q02722 mus musculus
c 7	14	0.8	585	1 ALU5_HUMAN	P39192 homo sapien
8	13	0.7	587	1 ALU2_HUMAN	P39189 homo sapien
c 9	13	0.7	587	1 ALU2_HUMAN	P39189 homo sapien
10	13	0.7	587	1 ALU3_HUMAN	P39190 homo sapien
c 11	13	0.7	587	1 ALU3_HUMAN	P39190 homo sapien
12	13	0.7	591	1 ALU1_HUMAN	P39188 homo sapien
c 13	13	0.7	591	1 ALU1_HUMAN	P39188 homo sapien
14	13	0.7	593	1 ALU6_HUMAN	P39193 homo sapien
c 15	13	0.7	593	1 ALU6_HUMAN	P39193 homo sapien
16	13	0.7	603	1 ALU4_HUMAN	P39191 homo sapien
c 17	13	0.7	603	1 ALU4_HUMAN	P39191 homo sapien
18	12	0.7	169	1 Y385_YEAST	P47152 saccharomyc

19	12	0.7	881	1 PKP2_HUMAN	Q99959 homo sapien
20	11	0.6	515	1 CRBA_DROME	P29747 drosophila
c 21	11	0.6	629	1 Z195_HUMAN	Q14628 homo sapien
c 22	11	0.6	1411	1 TCOF_HUMAN	Q13428 homo sapien
23	10	0.6	111	1 YG2C_YEAST	P53245 saccharomyc
c 24	10	0.6	167	1 YK20_YEAST	P36133 saccharomyc
c 25	10	0.6	465	1 SEP3_MOUSE	Q92155 mus musculus
c 26	10	0.6	741	1 RED1_HUMAN	P78563 homo sapien
c 27	9	0.5	130	1 YNO3_YEAST	P53908 saccharomyc
c 28	9	0.5	139	1 YED3_YEAST	P32633 saccharomyc
c 29	9	0.5	144	1 YHP5_YEAST	P38808 saccharomyc
c 30	9	0.5	222	1 YC13_KLEPN	Q48459 klebsiella
c 31	9	0.5	231	1 NKGC_MACMU	Q9mzk6 macaca mula
c 32	9	0.5	245	1 TM4B_HUMAN	Q9ukr8 homo sapien
c 33	9	0.5	253	1 ASP_PLAFA	P13825 plasmodium
c 34	9	0.5	331	1 IF2B_MOUSE	Q99145 mus musculus
c 35	9	0.5	333	1 IF2B_HUMAN	P20042 homo sapien
c 36	9	0.5	333	1 IF2B_RABIT	P41035 cryctolagus
c 37	9	0.5	418	1 YYY1_HUMAN	P49646 homo sapien
c 38	9	0.5	437	1 EF1H_XENLA	Q91375 xenopus lae
c 39	9	0.5	447	1 KBF3_HUMAN	Q04860 homo sapien
c 40	9	0.5	472	1 SX14_DROME	P40656 drosophila
c 41	9	0.5	478	1 AMP2_HUMAN	P50579 homo sapien
c 42	9	0.5	478	1 AMP2_MOUSE	O08663 mus musculus
c 43	9	0.5	478	1 AMP2_RAT	P38062 rattus norv
c 44	9	0.5	501	1 GYG2_HUMAN	Q15488 homo sapien
c 45	9	0.5	529	1 NOP5_HUMAN	Q972x3 homo sapien

#### ALIGNMENTS

RESULT 1  
ALU5\_HUMAN  
ID ALU5\_HUMAN STANDARD; PRT; 585 AA.  
AC P39192;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alu subfamily SC sequence contamination warning entry.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95021758; PubMed=7935834;  
RA Claverie J.-M., Makalowski W.;  
RT "Alu alert.";  
RL Nature 371:752-752(1994).  
RN [2]  
RP CONCEPT.  
RX MEDLINE=92241891; PubMed=1572661;  
RA Claverie J.-M.;  
RT "Identifying coding exons by similarity search: alu-derived and other  
RT potentially misleading protein sequences.";  
RL Genomics 12:838-841(1992).  
RN [3]  
RX ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=88333009; PubMed=3138422;  
RA Quentin Y.;  
RT "The Alu family developed through successive waves of fixation  
RT closely connected with primate lineage history.";  
RL J. Mol. Evol. 27:194-202(1988).  
RN [4]  
RX ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=91178815; PubMed=1706781;  
RA Jurka J., Milosavljevic A.;  
RT "Reconstruction and analysis of human Alu genes.";  
RL J. Mol. Evol. 32:105-121(1991).  
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU

CC REPEATS.  
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
CC ACID SEQUENCES.  
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
CC BEING REPORTED.  
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
CC CODING NUCLEOTIDE SEQUENCE.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL: U14571; -; NOT\_ANNOTATED\_CDS.

KW Hypothetical protein 95 FRAME-1.  
FT DOMAIN 1 95  
FT DOMAIN 99 193 FRAME-2.  
FT DOMAIN 197 291 FRAME-3.  
FT DOMAIN 295 389 FRAME-4.  
FT DOMAIN 393 487 FRAME-5.  
FT DOMAIN 491 585 FRAME-6.  
SQ SEQUENCE 585 AA; 63957 MW; 46E8C4F493650A7 CRC64;

Alignment Scores:  
Pred. No.: 9.09e-08 Length: 585  
Score: 16.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.92% Indels: 0  
DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU05\_HUMAN (1-585)

OY 5448 TGTATTATTAGATGGGTTTCACCATGTTGCCAGGCTGTCT 5495  
Db 446 CysilepHeSerArgAspGlyValSerProCysTrpProGlyTrpSer 461

RESULT 2

ID ALU05\_HUMAN STANDARD; PRT; 591 AA.  
AC P39155;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Alu subfamily SX sequence contamination warning entry.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95021758; PubMed=7935834;  
RA Claverie J.-M., Makalowski W.;

RT "Alu alert.";  
RL Nature 371:752-752(1994).  
RN [2]  
RP CONCEPT.  
RX MEDLINE=92241891; PubMed=1572661;  
RA Claverie J.-M.;  
RT "Identifying coding exons by similarity search: alu-derived and other  
RL potentially misleading protein sequences.";  
RL Genomics 12:838-841(1992).  
RN [3]  
RP FAMILIES CLASSIFICATION.  
RX MEDLINE=88333009; PubMed=3138422;  
RA Quentin Y.;  
RT "The Alu family developed through successive waves of fixation  
RL closely connected with primate lineage history.";  
RL J. Mol. Evol. 27:194-202(1988).  
RN [4]  
RP FAMILIES CLASSIFICATION.  
RX MEDLINE=91178815; PubMed=1706781;  
RA Jurka J., Milosavljevic A.;  
RT "Reconstruction and analysis of human Alu genes.";  
RL J. Mol. Evol. 32:105-121(1991).  
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
CC REPEATS.  
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
CC ACID SEQUENCES.  
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
CC BEING REPORTED.  
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
CC CODING NUCLEOTIDE SEQUENCE.

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

DR EMBL: U14574; -; NOT\_ANNOTATED\_CDS.

KW Hypothetical protein.  
FT DOMAIN 1 96 FRAME-1.  
FT DOMAIN 100 195 FRAME-2.  
FT DOMAIN 199 294 FRAME-3.  
FT DOMAIN 298 393 FRAME-4.  
FT DOMAIN 397 492 FRAME-5.  
FT DOMAIN 496 591 FRAME-6.  
SQ SEQUENCE 591 AA; 64395 MW; AC8154AD8A6B280 CRC64;

Alignment Scores:  
Pred. No.: 9.08e-08 Length: 591  
Score: 16.00 Matches: 16  
Percent Similarity: 100.00% Conservative: 0

```
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.92% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU8_HUMAN (1-591)
QY 5448 TGTATTTTACAGAGTGGGTTTCACCATGTTGGCCAGGTGCTCT 5495
Db 549 CysillePheSerArgAspGlyValSerProCysTrpProGlyTrpSer 564

RESULT 3
ALU8_HUMAN
ID ALU8_HUMAN STANDARD; PRT; 591 AA.
AC P39195;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Alu subfamily SX sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:103-121(1991).
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
REPEATS.
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
ACID SEQUENCES.
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
BEING REPORTED.
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
```

```
WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CODING NUCLEOTIDE SEQUENCE.
CC
CC
CC
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U14574; -; NOT_ANNOTATED_CDS.
KW Hypothetical protein.
FT DOMAIN 1 96 FRAME-1.
FT DOMAIN 100 195 FRAME-2.
FT DOMAIN 199 294 FRAME-3.
FT DOMAIN 298 393 FRAME-4.
FT DOMAIN 397 492 FRAME-5.
FT DOMAIN 496 591 FRAME-6.
SQ SEQUENCE 591 AA; 64395 MW; AC8154AD8A6BB280 CRC64;

Alignment Scores:
Pred. No.: 9.08e-08 Length: 591
Score: 16.00 Matches: 16
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.92% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU8_HUMAN (1-591)
QY 5494 GACCAGCTGCGCCACATGTCACACCCCTCTACTTAAATACAA 5447
Db 127 AspGlnProGlyGlnHisGlyClnThrProSerLeuLeuLysIleGln 142

RESULT 4
ALU7_HUMAN
ID ALU7_HUMAN STANDARD; PRT; 593 AA.
AC P39194;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily SQ sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL
```

RL J. Mol. Evol. 32:105-121(1991).  
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON, 'XX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 CC ACID SEQUENCES.  
 CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS.  
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF A  
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
 CC BEING REPORTED.  
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: U14573; ; NOT\_ANNOTATED\_CDS.  
 CC KW Hypothetical protein.  
 CC FT DOMAIN 1 97 FRAME-1.  
 CC FT DOMAIN 101 196 FRAME-2.  
 CC FT DOMAIN 200 295 FRAME-3.  
 CC FT DOMAIN 299 395 FRAME-4.  
 CC FT DOMAIN 399 494 FRAME-5.  
 CC FT DOMAIN 498 593 FRAME-6.  
 CC SQ SEQUENCE 593 AA; 64417 MW; 54A4F50F33A6089F CRC64;  
 CC  
 CC Alignment Scores:  
 CC Pred. No.: 9.07e-08 Length: 593  
 CC Score: 16.00 Matches: 16  
 CC Percent Similarity: 100.00% Conservative: 0  
 CC Best Local Similarity: 100.00% Mismatches: 0  
 CC Query Match: 0.92% Indels: 0  
 CC DB: 1 Gaps: 0  
 CC  
 CC US-09-966-880A-9 (1-5514) x ALU7\_HUMAN (1-593)  
 CC Qy 5448 TGATTTTGTAGATGGGTTTCACCATGTGGCCAGGTGTCT 5495  
 CC Db 353 CysilepSerargaspGlyValSerProCysTrpProGlyTrpSer 368  
 CC  
 CC RESULT 5  
 CC ALU7\_HUMAN  
 CC ID ALU7\_HUMAN STANDARD; PRT; 593 AA.  
 CC AC P39194;  
 CC DT 01-FEB-1995 (Rel. 31, Created)  
 CC DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Alu subfamily SQ sequence contamination warning entry.  
 CC OS Homo sapiens (human).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M., Makalowski W.;  
 RT "Alu alert.";  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie J.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 RT potentially misleading protein sequences.";  
 RL Genomics 12:838-841(1992).  
 RN [3]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=88333009; PubMed=3138422;  
 RA Quentin Y.;  
 RT "The Alu family developed through successive waves of fixation  
 RT closely connected with primate lineage history.";  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=9117815; PubMed=1706781;  
 RA Jurka J., Mikosavljevic A.;  
 RT "Reconstruction and analysis of human Alu genes.";  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON, 'XX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 CC ACID SEQUENCES.  
 CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS.  
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
 CC BEING REPORTED.  
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: U14573; ; NOT\_ANNOTATED\_CDS.  
 CC KW Hypothetical protein.  
 CC FT DOMAIN 1 97 FRAME-1.  
 CC FT DOMAIN 101 196 FRAME-2.  
 CC FT DOMAIN 200 295 FRAME-3.  
 CC FT DOMAIN 299 395 FRAME-4.  
 CC FT DOMAIN 399 494 FRAME-5.  
 CC FT DOMAIN 498 593 FRAME-6.  
 CC SQ SEQUENCE 593 AA; 64417 MW; 54A4F50F33A6089F CRC64;  
 CC  
 CC Alignment Scores:  
 CC Pred. No.: 9.07e-08 Length: 593  
 CC Score: 16.00 Matches: 16  
 CC Percent Similarity: 100.00% Conservative: 0  
 CC Best Local Similarity: 100.00% Mismatches: 0  
 CC Query Match: 0.92% Indels: 0  
 CC DB: 1 Gaps: 0  
 CC  
 CC US-09-966-880A-9 (1-5514) x ALU7\_HUMAN (1-593)  
 CC Qy 5448 TGATTTTGTAGATGGGTTTCACCATGTGGCCAGGTGTCT 5495  
 CC Db 353 CysilepSerargaspGlyValSerProCysTrpProGlyTrpSer 368  
 CC  
 CC RESULT 5  
 CC ALU7\_HUMAN  
 CC ID ALU7\_HUMAN STANDARD; PRT; 593 AA.  
 CC AC P39194;  
 CC DT 01-FEB-1995 (Rel. 31, Created)  
 CC DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 CC DE Alu subfamily SQ sequence contamination warning entry.  
 CC OS Homo sapiens (human).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.





KW Hypothetical protein.  
 FT DOMAIN 1 95  
 FT FRAME-1.  
 FT DOMAIN 99 193  
 FT FRAME-2.  
 FT DOMAIN 197 291  
 FT FRAME-3.  
 FT DOMAIN 295 389  
 FT FRAME-4.  
 FT DOMAIN 393 487  
 FT FRAME-5.  
 FT DOMAIN 491 585  
 FT FRAME-6.  
 SQ SEQUENCE 585 AA: 63957 MW; 46E8C4F493650A7 CRC64;  
 Alignment Scores:  
 Pred. No.: 1.24e-05  
 Score: 14.00 Length: 585  
 Percent Similarity: 100.00% Matches: 14  
 Best Local Similarity: 100.00% Conservative: 0  
 Query Match: 0.81% Mismatches: 0  
 DB: 1 Indels: 0  
 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU5\_HUMAN (1-585)  
 QY 5488 CTTGGCCAAATGTTGAACCCATCTCTACTAAATAACAA 5447  
 DB 225 ProGlyGlnHisGlyGlnThrProSerLeuLeuLysileGln 238

RESULT 8  
 ID ALU2\_HUMAN  
 AC P39189; STANDARD; PRT; 587 AA.  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Alu subfamily SB sequence contamination warning entry.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M.; Makalowski W.;  
 RT "Alu alert.";  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie J.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 RT potentially misleading protein sequences.";  
 RL Genomics 12:838-841(1992).  
 RN [3]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=88333009; PubMed=3138422;  
 RA Quentin Y.;  
 RT "The Alu family developed through successive waves of fixation  
 RT closely connected with primate lineage history.";  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J.; Milosavljevic A.;  
 RT "Reconstruction and analysis of human Alu genes.";  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 CC ACID SEQUENCES.  
 CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE

CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS.  
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES.  
 CC BEING REPORTED.  
 CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.

CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U14568; -; NOT\_ANNOTATED\_CDS.  
 KW Hypothetical protein.  
 FT DOMAIN 1 96 FRAME-1.  
 FT DOMAIN 100 194 FRAME-2.  
 FT DOMAIN 198 292 FRAME-3.  
 FT DOMAIN 296 391 FRAME-4.  
 FT DOMAIN 395 489 FRAME-5.  
 FT DOMAIN 493 587 FRAME-6.  
 SQ SEQUENCE 587 AA: 63703 MW; 3EAB3E3E3929203 CRC64;

Alignment Scores:  
 Pred. No.: 0.000145  
 Score: 13.00 Length: 587  
 Percent Similarity: 100.00% Matches: 13  
 Best Local Similarity: 100.00% Conservative: 0  
 Query Match: 0.75% Mismatches: 0  
 DB: 1 Indels: 0  
 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU2\_HUMAN (1-587)  
 QY 3568 TGGCTCAGCGCTGTAAATCCAGCACTCTGGAGGCTGAG 3606  
 DB 5 TrpLeuthrProValIleProAlaLeuTrpGluAlaGlu 17

RESULT 9  
 ID ALU2\_HUMAN  
 AC P39189; STANDARD; PRT; 587 AA.  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Alu subfamily SB sequence contamination warning entry.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M.; Makalowski W.;  
 RT "Alu alert.";  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie J.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 RT potentially misleading protein sequences.";

RL Genomics 12:838-841(1992).  
RN [3]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=88333009; PubMed=3138422;  
RA Quentin Y.;  
RT "The Alu family developed through successive waves of fixation  
RT closely connected with primate lineage history.";  
RL J. Mol. Evol. 27:194-202(1988).  
RN [4]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=91178815; PubMed=1706781;  
RA Jurka J., Milosavljevic A.;  
RT "Reconstruction and analysis of human Alu genes.";  
RL J. Mol. Evol. 32:105-121(1991).  
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
CC REPEATS.  
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
CC ACID SEQUENCES.  
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
CC BEING REPORTED.  
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
CC CODING NUCLEOTIDE SEQUENCE.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U14568; -: NOT\_ANNOTATED\_CDS.  
KW Hypothetical protein.  
FT DOMAIN 1 96 FRAME-1.  
FT DOMAIN 100 194 FRAME-2.  
FT DOMAIN 198 292 FRAME-3.  
FT DOMAIN 296 391 FRAME-4.  
FT DOMAIN 395 489 FRAME-5.  
FT DOMAIN 493 587 FRAME-6.  
SQ SEQUENCE 587 AA; 63703 MW; 3EAAB3E3E3929203 CRC64;  
  
Alignment Scores:  
Pred. No.: 0.000145 Length: 587  
Score: 13.00 Matches: 13  
Percent Similarity: 100.00% Conservativeness: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.75% Indels: 0  
DB: 1 Gaps: 0  
  
US-09-966-880A-9 (1-5514) x ALU2\_HUMAN (1-587)  
  
Qy 3605 TCAGCCTCCAGAGTGTGGATTACAGGCGTGAGCCAC 3567  
|||||

Db 474 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHis 486  
RESULT 10  
ALU3\_HUMAN  
ID ALU3\_HUMAN STANDARD; PRT; 587 AA.  
AC P39190;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alu subfamily SBI sequence contamination warning entry.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95021758; PubMed=7935834;  
RA Claverie J.-M., Makalowski W.;  
RT "Alu alert.";  
RL Nature 371:752-752(1994).  
RN [2]  
RP CONCEPT.  
RX MEDLINE=92241891; PubMed=1572661;  
RA Claverie J.-M.;  
RT "Identifying coding exons by similarity search: alu-derived and other  
RT potentially misleading protein sequences.";  
RL Genomics 12:838-841(1992).  
RN [3]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=88333009; PubMed=3138422;  
RA Quentin Y.;  
RT "The Alu family developed through successive waves of fixation  
RT closely connected with primate lineage history.";  
RL J. Mol. Evol. 27:194-202(1988).  
RN [4]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=91178815; PubMed=1706781;  
RA Jurka J., Milosavljevic A.;  
RT "Reconstruction and analysis of human Alu genes.";  
RL J. Mol. Evol. 32:105-121(1991).  
CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
CC REPEATS.  
CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
CC ACID SEQUENCES.  
CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
CC BEING REPORTED.  
CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
CC CODING NUCLEOTIDE SEQUENCE.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 CC EMBL; U14569; -: NOT\_ANNOTATED\_CDS.

CC KW Hypothetical protein. FRAME-1. 96  
 FT DOMAIN 1 100 194  
 FT DOMAIN 100 194  
 FT DOMAIN 198 292  
 FT DOMAIN 296 391  
 FT DOMAIN 395 489  
 FT DOMAIN 493 587  
 FT DOMAIN 587 587  
 SQ SEQUENCE 587 AA; 63573 MW; 85C4155726DEF235 CRC64;

Alignment Scores:  
 Pred. No.: 0.000145 Length: 587  
 Score: 13.00 Matches: 13  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 0.75% Indels: 0  
 DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU3\_HUMAN (1-587)

QY 3568 TGGCTCAGCCTGTAATCCAGCACTCTGGGAGGCTGAG 3606

Db 5 TrpLeuThrProValIleProAlaLeuTrpGluAlaGlu 17

RESULT 11

ID ALU3\_HUMAN STANDARD; PRT; 587 AA.

AC P39190;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alu subfamily SBI sequence contamination warning entry.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M., Makalowski W.;  
 RT "Alu alert.";  
 RL Nature 371:752-752(1994).  
 RN [2]

RP CONCEPT  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie J.-M.;

RT "Identifying coding exons by similarity search: alu-derived and other  
 RT potentially misleading protein sequences.";  
 RL Genomics 12:838-841(1992).  
 RN [3]

RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=86333009; PubMed=3138422;  
 RA Quentin Y.;

RT "The Alu family developed through successive waves of fixation  
 RT closely connected with primate lineage history.";  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]

RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J., Milosavljevic A.;

RT "Reconstruction and analysis of human Alu genes.";  
 RL J. Mol. Evol. 32:105-121(1991).  
 RN [5]

CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.

CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
 CC ACID SEQUENCES.

CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
 CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
 CC BEING REPORTED.

CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
 CC CODING NUCLEOTIDE SEQUENCE.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U14569; -: NOT\_ANNOTATED\_CDS.

DR KW Hypothetical protein. FRAME-1. 96  
 FT DOMAIN 1 100 194  
 FT DOMAIN 100 194  
 FT DOMAIN 198 292  
 FT DOMAIN 296 391  
 FT DOMAIN 395 489  
 FT DOMAIN 493 587  
 FT DOMAIN 587 587  
 SQ SEQUENCE 587 AA; 63573 MW; 85C4155726DEF235 CRC64;

Alignment Scores:  
 Pred. No.: 0.000145 Length: 587  
 Score: 13.00 Matches: 13  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 0.75% Indels: 0  
 DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU3\_HUMAN (1-587)

QY 3605 TCAGCTCCAGAGTGGCTGATTACAGCGTGAGCCAC 3567

Db 474 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHis 486

RESULT 12

ID ALU1\_HUMAN STANDARD; PRT; 591 AA.

AC P39188;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alu subfamily J sequence contamination warning entry.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M., Makalowski W.;  
 RT "Alu alert.";  
 RL Nature 371:752-752(1994).  
 RN [2]

```
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U14567; -; NOT_ANNOTATED_CDS.
FT DOMAIN 1 96
FT DOMAIN 100 195
FT DOMAIN 199 294
FT DOMAIN 298 393
FT DOMAIN 397 492
FT DOMAIN 496 591
FT SEQUENCE 591 AA; 63790 MW; 665D395735519D95 CRC64;
SQ
Alignment Scores:
Pred. No.: 0.000145 Length: 591
Score: 13.00 Matches: 13
Percent Similarity: 100.00% Conservatve: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.75% Indels: 0
DB: 1 Gaps: 0
```

```
US-09-966-880A-9 (1-5514) x ALU1_HUMAN (1-591)
QY 3568 TGGCTCAGCCCTGTATATCCAGCAGCTCTGGAGGCTGAG 3606
DB 5 TrpLeuThrProValIleProAlaLeuTrpGluAlaGlu 17
RESULT 13
ALU1_HUMAN STANDARD; PRT; 591 AA.
ID ALU1_HUMAN AC P39188;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Alu subfamily J sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752(1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;
RA Claverie J.-M.;
RT "Identifying coding exons by similarity search: alu-derived and other
RT potentially misleading protein sequences.";
RL Genomics 12:838-841(1992).
RN [3]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=88333009; PubMed=3138422;
RA Quentin Y.;
RT "The Alu family developed through successive waves of fixation
RT closely connected with primate lineage history.";
RL J. Mol. Evol. 27:194-202(1988).
RN [4]
RP ALU FAMILIES CLASSIFICATION.
RX MEDLINE=91178815; PubMed=1706781;
RA Jurka J., Milosavljevic A.;
RT "Reconstruction and analysis of human Alu genes.";
RL J. Mol. Evol. 32:105-121(1991).
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC REPEATS.
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
CC ACID SEQUENCES.
CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
CC BEING REPORTED.
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
CC CODING NUCLEOTIDE SEQUENCE.
```

-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; U14567; -; NOT\_ANNOTATED\_CDS.  
CC  
CC KW Hypothetical protein.  
CC FT DOMAIN 1 96 FRAME-1.  
CC FT DOMAIN 100 195 FRAME-2.  
CC FT DOMAIN 199 294 FRAME-3.  
CC FT DOMAIN 298 393 FRAME-4.  
CC FT DOMAIN 397 492 FRAME-5.  
CC FT DOMAIN 496 591 FRAME-6.  
CC SQ SEQUENCE 591 AA; 63790 MW; 665D395735519D95 CRC64;  
  
Alignment Scores:  
Pred. No.: 0.000145 Length: 591  
Score: 13.00 Matches: 13  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.75% Indels: 0  
DB: 1 Gaps: 0  
  
US-09-966-880A-9 (1-5514) x ALU1\_HUMAN (1-591)  
  
QY 3605 TCAGCTCCAGAGTCGTGGGATTCACAGCGTGACCCAC 3567  
DB 378 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHis 390  
  
RESULT 14  
ALU6\_HUMAN  
ID ALU6\_HUMAN STANDARD; PRT; 593 AA.  
AC P39193;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alu subfamily SP sequence contamination warning entry.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95021758; PubMed=7935834;  
RA Claverie J.-M., Makalowski W.;  
RT "Alu alert."  
RL Nature 371:752-752(1994).  
RN [2]  
RP CONCEPT.  
RX MEDLINE=92241891; PubMed=1572661;  
RA Claverie J.-M.;  
RT "Identifying coding exons by similarity search: alu-derived and other  
RT potentially misleading protein sequences.";  
RL Genomics 12:838-841(1992).  
RN [3]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=88333009; PubMed=3138422;  
RA Quentin Y.;  
RT "The Alu family developed through successive waves of fixation  
RT closely connected with primate lineage history.";  
RL J. Mol. Evol. 27:194-202(1988).  
RN [4]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=91178815; PubMed=1706781;  
RA Jurka J., Milosavljevic A.;  
RT "Reconstruction and analysis of human Alu genes.";  
RL J. Mol. Evol. 32:105-121(1991).  
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING

-----  
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
CC REPEATS.  
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
CC ACID SEQUENCES.  
CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
CC ACTIVELY TRANSCRIBED BY POL. III. NORMAL TRANSCRIPTS MAY CONTAIN  
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
CC BEING REPORTED.  
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
CC CODING NUCLEOTIDE SEQUENCE.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL; U14572; -; NOT\_ANNOTATED\_CDS.  
CC  
CC KW Hypothetical protein.  
CC FT DOMAIN 1 97 FRAME-1.  
CC FT DOMAIN 101 196 FRAME-2.  
CC FT DOMAIN 200 295 FRAME-3.  
CC FT DOMAIN 299 395 FRAME-4.  
CC FT DOMAIN 399 494 FRAME-5.  
CC FT DOMAIN 498 593 FRAME-6.  
CC SQ SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;  
  
Alignment Scores:  
Pred. No.: 0.000145 Length: 593  
Score: 13.00 Matches: 13  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.75% Indels: 0  
DB: 1 Gaps: 0  
  
US-09-966-880A-9 (1-5514) x ALU6\_HUMAN (1-593)  
  
QY 3568 TGGCTCAGCCCTGTAATCCAGCACTCTGGGAGGCTGAG 3606  
DB 5 TrpLeuThrProValIleProAlaLeuTrpGluAlaGlu 17  
  
RESULT 15  
ALU6\_HUMAN  
ID ALU6\_HUMAN STANDARD; PRT; 593 AA.  
AC P39193;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alu subfamily SP sequence contamination warning entry.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.

RX MEDLINE=95021758; PubMed=7935834;  
RA Claverie J.-M., Makalowski W.;  
RT "Alu alert."  
RL Nature 371:752-752(1994).  
RN [2]  
RP CONCEPT.  
RX MEDLINE=92241891; PubMed=1572661;  
RA Claverie J.-M.;  
RT "Identifying coding exons by similarity search: alu-derived and other  
RT potentially misleading protein sequences.";  
RL Genomics 12:838-841(1992).  
RN [3]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=88333009; PubMed=3138422;  
RA Quentin Y.;  
RT "The Alu family developed through successive waves of fixation  
RT closely connected with primate lineage history.";  
RL J. Mol. Evol. 27:194-202(1988).  
RN [4]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=91178815; PubMed=1706781;  
RA Jurka J., Milosavljevic A.;  
RT "Reconstruction and analysis of human Alu genes.";  
RL J. Mol. Evol. 32:103-121(1991).  
CC -|- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
CC REPEATS.  
CC -|- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
CC -|- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
CC ACID SEQUENCES.  
CC -|- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
CC BEING REPORTED.  
CC -|- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
CC CODING NUCLEOTIDE SEQUENCE.  
CC -----

This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

EMBL; U14572; ; NOT\_ANNOTATED\_CDS.  
DR Hypothetical protein.  
KW DOMAIN 1 97 FRAME-1.  
FT DOMAIN 101 196 FRAME-2.  
FT DOMAIN 200 295 FRAME-3.  
FT DOMAIN 299 395 FRAME-4.  
FT DOMAIN 399 494 FRAME-5.  
FT DOMAIN 498 593 FRAME-6.  
SQ SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;

Alignment Scores:

Pred. No.: 0.000145 Length: 593

Score: 13.00 Matches: 13  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.75% Indels: 0  
DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x ALU6\_HUMAN (1-593)

QY 3605 TCAGGCTCCAGAGTCTGGGATTACAGCGTGAGCCAC 3567

Db 479 SerAlaSerGlnSerAlaGlylleThrGlyValSerHis 491

Search completed: June 14, 2003, 19:19:23

Job time : 107.5 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 14, 2003, 18:57:34 ; Search time 332 Seconds  
(without alignments)  
6844.240 Million cell updates/sec

Title: US-09-966-880A-9

Perfect score: 1736

Sequence: 1 acagacgaatacatagtcca.....tcaaacctcgtacctcagag 5514

Scoring table: OLIGO  
Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0  
  
Searched: 671580 seqs, 206047115 residues

Word size: 1

Total number of hits satisfying chosen parameters: 1343044

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO\_SPOOL/US09966880/runat\_14062003\_175621\_11067/app\_query.fasta\_1.5703  
-DB=SPTREMBL\_21 -QFW=fastan -SUFFIX=olin2p.rspt -MINMATCH=0.1 -LOOPEXT=0  
-DOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09966880 -CGN\_1\_1\_587 -runat\_14062003\_175621\_11067 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGESQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -XGAPEXT=60 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rvirus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
c 1	22	1.3	449	4	Q9NX12	Q9nx12 homo sapien

c 2	18	1.0	101	11	Q9D443	Q9d443 mus musculu
c 3	18	1.0	104	5	Q9NLE3	Q9nle3 leishmania
c 4	18	1.0	188	5	Q9N8S9	Q9n8s9 trypanosoma
c 5	18	1.0	322	5	P90550	P90550 leishmania
c 6	18	1.0	413	12	Q95566	Q95566 human herpe
c 7	17	-1.0	102	4	Q8W188	Q8wy88 homo sapien
c 8	17	1.0	112	11	Q9D324	Q9d324 mus musculu
c 9	17	1.0	535	4	Q96EB1	Q96eb1 homo sapien
c 10	16	0.9	47	4	Q96CJ4	Q96cj4 homo sapien
c 11	16	0.9	50	4	Q96EM6	Q96em6 homo sapien
c 12	16	0.9	61	4	Q9BV30	Q9bv30 homo sapien
c 13	16	0.9	73	4	Q9GAN9	Q9gan9 homo sapien
c 14	16	0.9	121	4	Q9HAJ0	Q9haj0 homo sapien
c 15	16	0.9	160	4	Q9H654	Q9h654 homo sapien
c 16	16	0.9	175	11	Q9D399	Q9d399 mus musculu
c 17	15	0.9	40	4	Q8W73	Q8w73 homo sapien
c 18	15	0.9	46	4	Q8TBW3	Q8tbw3 homo sapien
c 19	15	0.9	46	4	Q9BVC1	Q9bvc1 homo sapien
c 20	15	0.9	50	11	Q62710	Q62710 rattus norv
c 21	15	0.9	103	11	Q9CXB3	Q9cxb3 mus musculu
c 22	15	0.9	118	4	Q9H387	Q9h387 homo sapien
c 23	15	0.9	123	4	Q9HAD8	Q9had8 homo sapien
c 24	15	0.9	127	11	Q9D0V0	Q9d0v0 mus musculu
c 25	15	0.9	133	4	Q96JR5	Q96jrs homo sapien
c 26	15	0.9	143	11	Q64150	Q64150 mus musculu
c 27	15	0.9	257	4	Q9H7F8	Q9h7f8 homo sapien
c 28	14	0.8	50	11	Q8R193	Q8r193 mus musculu
c 29	14	0.8	72	4	Q9UI79	Q9ui79 homo sapien
c 30	14	0.8	128	11	Q91WM8	Q91wm8 mus musculu
c 31	14	0.8	139	4	Q9GZ25	Q9gz25 homo sapien
c 32	14	0.8	207	11	Q9D3D2	Q9d3d2 mus musculu
c 33	14	0.8	231	4	Q9NR08	Q9nr08 homo sapien
c 34	14	0.8	318	4	Q96QL7	Q96ql7 homo sapien
c 35	14	0.8	375	4	Q60448	Q60448 homo sapien
c 36	14	0.8	423	4	Q9NWF0	Q9nwf0 homo sapien
c 37	14	0.8	461	5	Q9N7X3	Q9n7x3 leishmania
c 38	13	0.7	29	4	Q96IR5	Q96ir5 homo sapien
c 39	13	0.7	90	4	Q96IG1	Q96ig1 homo sapien
c 40	13	0.7	130	4	Q9HBS7	Q9hbs7 homo sapien
c 41	13	0.7	174	4	Q9H926	Q9h926 homo sapien
c 42	13	0.7	208	4	Q9GQ97	Q9gq97 homo sapien
c 43	13	0.7	232	4	Q9H5R3	Q9h5r3 homo sapien
c 44	13	0.7	296	4	Q8WWD5	Q8wwd5 homo sapien
c 45	13	0.7	296	7	Q29890	Q29890 homo sapien

ALIGNMENTS

RESULT 1

Q9NX12	PRELIMINARY;	PRT;	449 AA.
ID	Q9NX12		
AC	Q9NX12:		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DE	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)		
DE	CDNA FLJ20234 fis, clone COLF5673.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=COLON MUCOSA;		
RA	Watanabe K., Kumeagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,		
RA	Suzuki Y., Ohtsubashi M., Nishi T., Shibahara T., Tanaka T.,		
RA	Nakamura Y., Isogai T., Sugano S.;		
RT	"NEO human cDNA sequencing project."		
RL	Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AK000241; BAA91028.1;		
DR	InterPro; IPR001849; PH.		
DR	InterPro; IPR000980; SH2.		
DR	Pfam; PF00017; SH2; 1.		
DR	SMART; SM00233; PH; 1.		







[illegible]

```

US-09-966-880A-9 (1-5514) x Q96AN9 (1-73)
QY      5448 TGTATTTTACTAGATGGGTTTCACCATGTTGGCCAGCGTGCTC 5495
Db      52 CysIlePheSerArgAspGlyValSerProCysTrpProGlyTrpSer 67

RESULT 14
Q9HAJ0
ID      O9HAJ0          PRELIMINARY;          PRT;       121 AA.
AC      Q9HAJ0;
CD
DT      01-MAR-2001 (TrEMBLrel. 16, Created)
DT      01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE      Hypothetical 13.7 kDa protein.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN      [1] TaxID=9606;
RP      SEQUENCE FROM N.A.
RC      TISSUE=EMBRYO;
RA      Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA      Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA      WagaTsuna M., Hosofri T., Kakui Y., Kodaira H., Kondo H., Sugawara
RA      Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi
RA      Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA      Yamamoto J., Nakamatsu A., Nakamura Y., Nagahari K., Masuo Y.,
RA      Nedomiya K., Iwayanagi T.;
RT      "NEBO human cDNA sequencing project.";
RL      Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AK021618; BABI3856.1;
SQ      SEQUENCE 121 AA; 13691 MW; 265372CF6CBF8E56 CRC64;

Alignment Scores:
Pred. No.:           9,74e-07              Length:         121
Score:               16.00                 Matches:         16
Percent Similarity: 100.00%                Conservative:    0
Best Local Similarity: 100.00%             Mismatches:     0
Query Match:        0.92%                  Indels:         0
DB:                 4                      Gaps:           0

US-09-966-880A-9 (1-5514) x Q9HAJ0 (1-121)
QY      2890 TGTGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG 2843
Db      72 CysValCysValCysValCysValCysValCysValCysValCysValCysVal 87

RESULT 15
Q9H654
ID      Q9H654          PRELIMINARY;          PRT;       160 AA.
AC      Q9H654;
CD
DT      01-MAR-2001 (TrEMBLrel. 16, Created)
DT      01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE      CDNA: FLJ22596 fis, clone HS103808.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN      [1] NCBI_TaxID=9606;
RP      SEQUENCE FROM N.A.
RC      TISSUE=SMALL INTESTINE;
RA      Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA      Suzuki Y., Ohbayashi M., Nishi T., Shibahara T., Tanaka T.,
RA      Nakamura Y., Isoigai T., Sugano S.;
RT      "NEBO human cDNA sequencing project.";
RL      Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AK026249; BABI5412.1;
SQ      SEQUENCE 160 AA; 16738 MW; FB8B34293CE827E7 CRC64;

Alignment Scores:
Pred. No.:           9.38e-07              Length:         160
Score:               16.00                 Matches:         16

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - protein search, using frame\_plus\_n2p model

Run on: June 14, 2003, 19:09:34 ; Search time 52.5 Seconds  
(without alignments)  
6180.495 Million cell updates/sec

Title: US-09-966-880A-9  
Perfect score: 1736  
Sequence: 1 acagacgaatacatagctgtccca.....tcaaaactcctgacctcagag 5514

Scoring table: OLIGO

Xgapop 60.0 , Xgapext 60.0  
Ygapop 60.0 , Ygapext 60.0  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Word size: 1

Total number of hits satisfying chosen parameters: 438908

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=xl  
-Q/cgn2\_1/USPTO\_Spool/US09966880/runat\_14062003\_175623\_11106/app\_query.fasta\_1.5703  
-DB=Issued\_Patents\_AA -QMT=fastan -SUFFIX=olin2p.ra -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=oligo -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=quality -THR\_MIN=1 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pto -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09966880 -CGN\_1\_1\_83@runat\_14062003\_175623\_11106 -NCRU=6 -ICPU=3  
-NO\_WMAP -LARGESQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=60 -FGAPEXT=60 -FGAPOB=6  
-FGAPEXT=7 -YGAPOP=60 -YGAPEXT=60 -DELOP=6 -DELEXT=7

Database :

Issued Patents\_AA:  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	14	0.8	90	4	US-09-227-357-171
c 2	14	0.8	96	4	US-09-605-785-827
c 3	14	0.8	224	4	US-09-605-785-825
c 4	14	0.8	375	2	US-08-454-557C-121
c 5	14	0.8	375	2	US-08-340-426D-121
c 6	14	0.8	375	2	US-08-450-673C-121
c 7	14	0.8	410	1	US-08-153-848-7
c 8	14	0.8	410	3	US-09-299-843A-7
c 9	14	0.8	410	4	US-09-088-337B-7
c 10	14	0.8	410	5	PCT-US93-11153-7
c 11	14	0.8	1079	3	US-09-058-489-22
c 12	14	0.8	1261	4	US-09-605-785-538

13	14	0.8	1261	4	US-09-439-313-538	Sequence 538, App
14	13	0.7	65	4	US-09-227-357-538	Sequence 538, App
c 15	13	0.7	70	4	US-09-188-930-131	Sequence 131, App
16	12	0.7	90	4	US-09-227-357-171	Sequence 171, App
17	12	0.7	122	4	US-09-227-357-285	Sequence 285, App
c 18	12	0.7	397	5	PCT-US95-17111A-121	Sequence 121, App
c 19	12	0.7	605	4	US-09-440-936-2	Sequence 2, Appli
20	11	0.6	52	4	US-09-227-357-537	Sequence 537, App
21	11	0.6	521	2	US-08-721-684C-2	Sequence 2, Appli
22	11	0.6	521	2	US-09-005-970-2	Sequence 2, Appli
23	11	0.6	521	4	US-09-407-715-2	Sequence 2, Appli
c 24	11	0.6	1079	3	US-09-058-489-22	Sequence 22, Appl
c 25	10	0.6	19	4	US-09-463-238-30	Sequence 30, Appl
c 26	10	0.6	32	4	US-09-288-143-171	Sequence 171, App
27	10	0.6	35	4	US-09-288-143-170	Sequence 170, App
c 28	10	0.6	36	1	US-07-948-357-12	Sequence 12, Appl
c 29	10	0.6	36	1	US-07-948-357-13	Sequence 13, Appl
c 30	10	0.6	36	2	US-08-450-417-12	Sequence 12, Appl
c 31	10	0.6	36	2	US-08-450-417-13	Sequence 13, Appl
c 32	10	0.6	36	3	US-08-449-741-12	Sequence 12, Appl
c 33	10	0.6	36	3	US-08-449-741-13	Sequence 13, Appl
c 34	10	0.6	36	3	US-08-449-754-12	Sequence 12, Appl
c 35	10	0.6	36	3	US-08-449-754-13	Sequence 13, Appl
c 36	10	0.6	36	4	US-09-465-646-12	Sequence 12, Appl
c 37	10	0.6	36	4	US-09-465-646-13	Sequence 13, Appl
38	10	0.6	49	4	US-09-288-143-93	Sequence 93, Appl
39	10	0.6	56	4	US-09-227-357-577	Sequence 577, App
40	10	0.6	80	4	US-09-227-357-249	Sequence 249, App
c 41	10	0.6	99	4	US-09-288-143-168	Sequence 168, App
c 42	10	0.6	99	4	US-09-288-143-168	Sequence 168, App
c 43	10	0.6	182	1	US-08-226-264-28	Sequence 28, Appl
c 44	10	0.6	312	4	US-09-254-465A-9	Sequence 9, Appli
45	10	0.6	388	4	US-09-265-630-11	Sequence 11, Appl

ALIGNMENTS

RESULT 1  
US-09-227-357-171  
; Sequence 171, Application US/09227357  
; Patent No. 6342581  
; GENERAL INFORMATION:  
; APPLICANT: Fischer et al.  
; TITLE OF INVENTION: 123 Human Secreted Proteins  
; FILE REFERENCE: P2010P1  
; CURRENT APPLICATION NUMBER: US/09/227,357  
; EARLIER FILING DATE: 1999-01-08  
; EARLIER APPLICATION NUMBER: PCT/US98/13684  
; EARLIER FILING DATE: 1998-07-07  
; EARLIER APPLICATION NUMBER: 60/051,926  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,793  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,925  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,929  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,803  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,732  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,931  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,932  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,916  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,930  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,918  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,920  
; EARLIER FILING DATE: 1997-07-08

```

; EARLIER APPLICATION NUMBER: 60/052,733
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,795
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,919
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,928
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/055,722
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,723
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,948
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,949
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,953
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,950
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,947
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,964
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/056,360
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,684
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,984
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/055,954
; EARLIER FILING DATE: 1997-08-18
; EARLIER APPLICATION NUMBER: 60/058,785
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,664
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,660
; EARLIER FILING DATE: 1997-09-12
; EARLIER APPLICATION NUMBER: 60/058,661
; EARLIER FILING DATE: 1997-09-12
; NUMBER OF SEQ ID NOS: 672
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 171
; LENGTH: 90
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (90)
; OTHER INFORMATION: Xaa equals stop translation
US-09-227-357-171

Alignment Scores:
Pred. No.: 0.000128 Length: 90
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 4 Gaps: 0

US-09-966-880a-9 (1-5514) x US-09-227-357-171 (1-90)
QY 3605 TCAGCCTCCAGAGTGTGGGATTACAGCGTGACCCAC 3564
Db 36 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHis 49

RESULT 2
US-09-605-785-827
; Sequence 827, Application us/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kaos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 827
; LENGTH: 96
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-605-785-827

Alignment Scores:
Pred. No.: 0.000127 Length: 96
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 4 Gaps: 0

US-09-966-880a-9 (1-5514) x US-09-605-785-827 (1-96)
QY 5462 AGATGGGTTTCACCATGTGGCCAGCGTGTCTCAAACTCC 5503
Db 75 ArgtrpGlyPheThrMetLeuAlaArgLeuValSerAsnSer 88

RESULT 3
US-09-605-785-825
; Sequence 825, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kaos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 825
; LENGTH: 224

```

;  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-605-785-825

Alignment Scores:  
Pred. No.: 0.000118 Length: 224  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.81% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-605-785-825 (1-224)

QY 5462 AGATGGGTTTCCACCATGTGCGAGGTGCTCAAACTCC 5503  
|||||  
Db 211 ArgTrpGlyPheThrMetLeuAlaArgLeuValSerAsnSer 224  
|||||

RESULT 4

US-08-454-557C-121  
; Sequence 121, Application US/08454557C  
; Patent No. 5830670  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/454,557C  
FILING DATE: 30-MAY-1995  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Ludwig, Steven R.  
REGISTRATION NUMBER: 36,203  
REFERENCE/DOCKET NUMBER: 0609.3840003  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 121:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 375 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-454-557C-121

Alignment Scores:  
Pred. No.: 0.000112 Length: 375  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.81% Indels: 0  
DB: 2 Gaps: 0

US-09-966-880A-9 (1-5514) x US-08-454-557C-121 (1-375)

QY 3605 TCAGCTCCAGAGTGGGATTACAGCGGTGAGCCACCAC 3564  
|||||  
Db 279 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHis 292  
|||||

Alignment Scores:

Pred. No.: 0.000112 Length: 375  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.81% Indels: 0  
DB: 2 Gaps: 0

US-09-966-880A-9 (1-5514) x US-08-454-557C-121 (1-375)

QY 3605 TCAGCTCCAGAGTGGGATTACAGCGGTGAGCCACCAC 3564  
|||||  
Db 279 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHis 292  
|||||

RESULT 5

US-08-340-426D-121  
; Sequence 121, Application US/08340426D  
; Patent No. 5948634  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/340,426D  
FILING DATE: 14-NOV-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Ludwig, Steven R.  
REGISTRATION NUMBER: 36,203  
REFERENCE/DOCKET NUMBER: 0609.3840002  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 121:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 375 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-340-426D-121

Alignment Scores:  
Pred. No.: 0.000112 Length: 375  
Score: 14.00 Matches: 14  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 0.81% Indels: 0  
DB: 2 Gaps: 0

US-09-966-880A-9 (1-5514) x US-08-340-426D-121 (1-375)

QY 3605 TCAGCTCCAGAGTGGGATTACAGCGGTGAGCCACCAC 3564  
|||||  
Db 279 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisHis 292  
|||||

RESULT 6

US-08-450-673C-121  
; Sequence 121, Application US/08450673C  
; Patent No. 5948888  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection  
; TITLE OF INVENTION: of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,673C
; FILING DATE: 30-MAY-1995
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Ludwig, Steven R.
; REGISTRATION NUMBER: 36,203
; REFERENCE/DOCKET NUMBER: 0609.3840004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 121:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 375 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-450-673C-121

Alignment Scores:
Pred. No.: 0.000112 Length: 375
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 2 Gaps: 0

US-09-966-880A-9 (1-5514) x US-08-450-673C-121 (1-375)

QY 3605 TCAGCTCCAGAGTGCTGGGATTACAGCGGTGAGCCAC 3564
|||||
Db 279 SerAlaSerGlnSerIaGlylleThrGlyValSerHisHis 292

RESULT 7
US-08-153-848-7
; Sequence 7, Application US/08153848
; Patent No. 5759804
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: No. 5759804el Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Bicknell
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/153,848
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5759804and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31794
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
```

```
; TELEFAX: (312) 474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 410 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-153-848-7

Alignment Scores:
Pred. No.: 0.000111 Length: 410
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 1 Gaps: 0

US-09-966-880A-9 (1-5514) x US-08-153-848-7 (1-410)

QY 3291 ATGATTGCACCACTGCACCTCCAGCCTGGGTAAACAGGCTGAGA 3250
|||||
Db 1 MetIleAlaProLeuHisSerSerLeuGlyAsnArgValArg 14

RESULT 8
US-09-299-843A-7
; Sequence 7, Application US/09299843A
; Patent No. 6107475
; GENERAL INFORMATION:
; APPLICANT: Godiska, Ronald
; APPLICANT: Gray, Patrick W.
; APPLICANT: Schweikart, Vicki L.
; TITLE OF INVENTION: No. 6107475el Seven Transmembrane Receptors
; NUMBER OF SEQUENCES: 66
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESSEE: Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/299,843A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/088,337
; FILING DATE: 01-JUN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/153,848
; FILING DATE: 17-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/977,452
; FILING DATE: 17-NOV-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Jill E. Uhl
; REGISTRATION NUMBER: 43,213
; REFERENCE/DOCKET NUMBER: 27866/32059B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 410 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
```

us-09-966-880a-9-9.olin2p.ra1

Fri Jun 20 09:21:58 2003

MOLECULE TYPE: protein  
US-09-299-843A-7

Alignment Scores: 0.000111 Length: 410  
Pred. No.: 14.00 Matches: 14  
Score: 100.00% Conservativity: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 0.81% Gaps: 0  
DB:

US-09-966-880A-9 (1-5514) x US-09-299-843A-7 (1-410)

QY 3291 ATGATTGCACCTGCACCTCCAGCTGGTAAACAGGGTGAGA 3250  
|||||  
Db 1 MetileAlaProleuHisSerLeuGlyAsnArgValArg 14

RESULT 9

US-09-088-337B-7  
; Sequence 7, Application US/09088337B  
; Patent No. 6348574

GENERAL INFORMATION:  
APPLICANT: Godiska, Ronald  
Gray, Patrick W.  
Schweikart, Vicki L.

TITLE OF INVENTION: No. 6348574el Seven Transmembrane Receptors

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/088.337B

FILING DATE: 01-Jun-1998

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/153,848

FILING DATE: 17-NOV-1993

APPLICATION NUMBER: US 07/977,452

FILING DATE: 17-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: No. 6348574and, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 31794

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 410 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 7:

US-09-088-337B-7

Alignment Scores: 0.000111 Length: 410  
Pred. No.: 14.00 Matches: 14  
Score: 100.00% Conservativity: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 0.81% Gaps: 0  
DB:

US-09-966-880A-9 (1-5514) x US-09-088-337B-7 (1-410)

QY 3291 ATGATTGCACCTGCACCTCCAGCTGGTAAACAGGGTGAGA 3250  
|||||  
Db 1 MetileAlaProleuHisSerLeuGlyAsnArgValArg 14

RESULT 10

PCT-US93-11153-7

; Sequence 7, Application PC/TUS9311153

GENERAL INFORMATION:

APPLICANT: Godiska, Ronald

Gray, Patrick W.

Schweikart, Vicki L.

TITLE OF INVENTION: Novel Seven Transmembrane Receptors

NUMBER OF SEQUENCES: 64

CORRESPONDENCE ADDRESS:

ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &

STREET: 6300 Sears Tower, 233 South Wacker Drive

CITY: Chicago

STATE: Illinois

COUNTRY: USA

ZIP: 60606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/11153

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/977,452

FILING DATE: 17-NOV-1992

ATTORNEY/AGENT INFORMATION:

NAME: Noland, Greta E.

REGISTRATION NUMBER: 35,302

REFERENCE/DOCKET NUMBER: 31794

TELECOMMUNICATION INFORMATION:

TELEPHONE: (312) 474-6300

TELEFAX: (312) 474-0448

TELEX: 25-3856

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 410 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

PCT-US93-11153-7

Alignment Scores: 0.000111 Length: 410

Pred. No.: 14.00 Matches: 14

Score: 100.00% Conservativity: 0

Percent Similarity: 100.00% Mismatches: 0

Best Local Similarity: 100.00% Indels: 0

Query Match: 0.81% Gaps: 0

DB:

US-09-966-880A-9 (1-5514) x PCT-US93-11153-7 (1-410)

QY 3291 ATGATTGCACCTGCACCTCCAGCTGGTAAACAGGGTGAGA 3250

|||||

Db 1 MetileAlaProleuHisSerLeuGlyAsnArgValArg 14

RESULT 11

US-09-058-489-22

; Sequence 22, Application US/09058489

; Patent No. 6103886

GENERAL INFORMATION:

APPLICANT: Whitehead Institute for Biomedical Research

APPLICANT: Lahn, Bruce

```

; APPLICANT: Page, David
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of
; FILE REFERENCE: WH197-08pA
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Human
US-09-058-489-22

Alignment Scores:
Pred. No.: 0.000102 Length: 1079
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 3 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-058-489-22 (1-1079)
QY 3605 TCAGCTCCAGAGTCTGGGATTACAGCGGTGAGCCAC 3564
Db 1064 SerAtaSerGlnSerAlaGlyIleThrGlyValSerHis 1077

RESULT 12
US-09-605-785-538
; Sequence 538, Application US/09605785
; Patent No. 6321716
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Jiang, Yuqi
; APPLICANT: Henderson, Robert A.
; APPLICANT: Kalos, Michael D.
; APPLICANT: Fanger, Gary R.
; APPLICANT: Retter, Marc W.
; APPLICANT: Stolk, John A.
; APPLICANT: Day, Craig H.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Carter, Darrick
; APPLICANT: Li, Samuel
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Hepler, William
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.427C16
; CURRENT APPLICATION NUMBER: US/09/605,785
; CURRENT FILING DATE: 2000-06-27
; NUMBER OF SEQ ID NOS: 835
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 538
; LENGTH: 1261
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-605-785-538

Alignment Scores:
Pred. No.: 0.0001 Length: 1261
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 4 Gaps: 0

```

```

US-09-966-880A-9 (1-5514) x US-09-605-785-538 (1-1261)
QY 5462 AGATGGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCC 5503
Db 1248 ArgTrpGlyPheThrMetLeuAlaArgLeuValSerAsnSer 1261

RESULT 13
US-09-439-313-538
; Sequence 538, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqi
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 538
; LENGTH: 1261
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-439-313-538

Alignment Scores:
Pred. No.: 0.0001 Length: 1261
Score: 14.00 Matches: 14
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 0.81% Indels: 0
DB: 4 Gaps: 0

US-09-966-880A-9 (1-5514) x US-09-439-313-538 (1-1261)
QY 5462 AGATGGGGTTTCACCATGTTGCCAGGCTGGTCTCAAACTCC 5503
Db 1248 ArgTrpGlyPheThrMetLeuAlaArgLeuValSerAsnSer 1261

RESULT 14
US-09-227-357-538
; Sequence 538, Application US/09227357
; Patent No. 6342581
; GENERAL INFORMATION:
; APPLICANT: Fischer et al.
; TITLE OF INVENTION: 123 Human Secreted Proteins
; FILE REFERENCE: P2010P1
; CURRENT APPLICATION NUMBER: US/09/227,357
; CURRENT FILING DATE: 1999-01-08
; EARLIER APPLICATION NUMBER: PCT/US98/13684
; EARLIER FILING DATE: 1998-07-07
; EARLIER APPLICATION NUMBER: 60/051,926
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,793
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,925
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/051,929
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,803
; EARLIER FILING DATE: 1997-07-08
; EARLIER APPLICATION NUMBER: 60/052,732
; EARLIER FILING DATE: 1997-07-08

```

Search completed: June 14, 2003, 19:37:45  
Job time : 90.5 secs

	EARLIER APPLICATION NUMBER: 60/051,931	
	EARLIER FILING DATE: 1997-07-08	
	EARLIER APPLICATION NUMBER: 60/051,932	
	EARLIER FILING DATE: 1997-07-08	
	EARLIER APPLICATION NUMBER: 60/051,916	
	EARLIER FILING DATE: 1997-07-08	
	EARLIER APPLICATION NUMBER: 60/051,930	
	EARLIER FILING DATE: 1997-07-08	
	EARLIER APPLICATION NUMBER: 60/051,918	
	EARLIER FILING DATE: 1997-07-08	
	EARLIER APPLICATION NUMBER: 60/051,920	
	EARLIER FILING DATE: 1997-07-08	
	EARLIER APPLICATION NUMBER: 60/052,733	
	EARLIER FILING DATE: 1997-07-08	
	EARLIER APPLICATION NUMBER: 60/052,795	
	EARLIER FILING DATE: 1997-07-08	
	EARLIER APPLICATION NUMBER: 60/051,919	
	EARLIER FILING DATE: 1997-07-08	
	EARLIER APPLICATION NUMBER: 60/051,928	
	EARLIER FILING DATE: 1997-07-08	
	EARLIER APPLICATION NUMBER: 60/055,722	
	EARLIER FILING DATE: 1997-08-18	
	EARLIER APPLICATION NUMBER: 60/055,723	
	EARLIER FILING DATE: 1997-08-18	
	EARLIER APPLICATION NUMBER: 60/055,948	
	EARLIER FILING DATE: 1997-08-18	
	EARLIER APPLICATION NUMBER: 60/055,949	
	EARLIER FILING DATE: 1997-08-18	
	EARLIER APPLICATION NUMBER: 60/055,953	
	EARLIER FILING DATE: 1997-08-18	
	EARLIER APPLICATION NUMBER: 60/055,950	
	EARLIER FILING DATE: 1997-08-18	
	EARLIER APPLICATION NUMBER: 60/055,947	
	EARLIER FILING DATE: 1997-08-18	
	EARLIER APPLICATION NUMBER: 60/055,964	
	EARLIER FILING DATE: 1997-08-18	
	EARLIER APPLICATION NUMBER: 60/056,360	
	EARLIER FILING DATE: 1997-08-18	
	EARLIER APPLICATION NUMBER: 60/055,684	
	EARLIER FILING DATE: 1997-08-18	
	EARLIER APPLICATION NUMBER: 60/055,984	
	EARLIER FILING DATE: 1997-08-18	
	EARLIER APPLICATION NUMBER: 60/055,954	
	EARLIER FILING DATE: 1997-08-18	
	EARLIER APPLICATION NUMBER: 60/058,785	
	EARLIER FILING DATE: 1997-09-12	
	EARLIER APPLICATION NUMBER: 60/058,664	
	EARLIER FILING DATE: 1997-09-12	
	EARLIER APPLICATION NUMBER: 60/058,660	
	EARLIER FILING DATE: 1997-09-12	
	EARLIER APPLICATION NUMBER: 60/058,661	
	EARLIER FILING DATE: 1997-09-12	
	NUMBER OF SEQ ID NOS: 672	
	SOFTWARE: PatentIn Ver. 2.0	
	SEQ ID NO 536	
	LENGTH: 65	
	TYPE: PRT	
	ORGANISM: Homo sapiens	
	US-09-227-357-538	
	Alignment Scores:	0.00118
	Pred. No.:	13.00
	Score:	Conservative: 0
	Percent Similarity:	100.00%
	Best Local Similarity:	100.00%
	Query Match:	0.75%
	Gaps:	4
	DB:	
	US-09-966-880A-9 (1-5514) x US-09-227-357-538 (1-65)	
	65 Length:	
	Matches:	13
	Conservative:	0
	Mismatches:	0
	Indels:	0
	Gaps:	0
	3577 CCTGTAATCCAGCACTCTCGGAGCGTGGGTGGGA 3615	
	QY	
	Db 43 ProValIleProAlaLeuTrpGluAlaGluValGlyCly 55	





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 20:16:19 ; Search time 576.918 seconds  
(without alignments)  
11000.051 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 2818  
Sequence: 1 agagaacatcattatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002: \*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT: \*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT: \*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT: \*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT: \*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT: \*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT: \*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT: \*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT: \*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT: \*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT: \*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT: \*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT: \*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT: \*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT: \*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT: \*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT: \*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT: \*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT: \*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT: \*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT: \*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT: \*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT: \*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT: \*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2818	100.0	2818	21 AAC55312	Human activation-i
2	2174.6	77.2	6564	21 AAC55314	Human activation-i
3	2174.6	77.2	11204	21 AAC55339	Human activation-i
4	2172	77.1	2172	21 AAC55313	Human activation-i
5	1489.2	52.8	1665	22 AAK81088	Human immune/haema
6	603.4	21.4	2440	21 AAC55307	Mouse activation-i
7	574	20.4	574	22 AAK81089	Human immune/haema
8	429	15.2	429	22 AAK61819	Human immune/haema
9	295.2	10.5	13862	22 ABA08208	Human ovarian and

10	295.2	10.5	13862	22	AAL02789	Human reproductive
11	295.2	10.5	13862	22	AAL07516	Human reproductive
12	292.6	10.4	7809	22	ABA15724	Human nervous syst
13	292.6	10.4	7809	22	AAL03406	Human reproductive
14	292.4	10.4	30393	22	AAK67239	Human immune/haema
15	291	10.3	25012	22	ABA15431	Human nervous syst
16	286.2	10.2	37314	22	AAK71358	Human immune/haema
17	286.2	10.2	172570	24	ABQ88207	Human osteoblast d
18	284.4	10.1	32249	22	AAL04676	Human reproductive
19	284.4	10.1	32249	23	ABL97583	Human testicular a
20	284.4	10.1	129722	24	ABQ88117	Human osteoblast d
21	283.4	10.1	1792	24	ABA96614	Human alpha interf
22	283.2	10.0	22428	22	AAS41759	Genomic sequence #
23	282.4	10.0	13919	24	ABK86218	DNA encoding AIP-1
24	282.4	10.0	13919	24	ABK86220	AIP-1/FLASH promot
25	282.4	10.0	13919	24	ABK86221	AIP-1/FLASH promot
26	282.4	10.0	21404	24	ABK86229	AIP-1/FLASH promot
27	281.4	10.0	65608	24	ABL62910	Breast cancer rela
28	281.4	10.0	65608	24	ABL64414	Oesophagus cancer
29	281.4	10.0	65608	24	ABL67668	Genomic sequence #
30	281	10.0	23452	22	AAS42122	Human musculoskele
31	280.2	9.9	11821	22	AAL36492	Human CDNA differe
32	280.2	9.9	122888	24	ABK83569	Human proto-oncoge
33	280	9.9	15297	24	ABA94501	Genomic sequence #
34	279.6	9.9	23457	22	AAS42121	Genomic sequence #
35	279.6	9.9	23458	22	AAS42120	Human CDNA differe
36	278.4	9.9	12263	24	ABK84514	Human immune/haema
37	278.4	9.9	21596	22	AAK71582	Human reproductive
38	278.2	9.9	32204	22	AAL05849	Human testicular a
39	278.2	9.9	32204	23	ABL98413	Human osteoblast d
40	277.6	9.9	86080	24	ABQ88164	Human CDNA differe
41	277.6	9.9	86080	24	ABK83561	Human digestive sy
42	277.2	9.8	5253	22	AAK89859	Human digestive sy
43	277.2	9.8	5257	22	AAK89860	Human immune/haema
44	277	9.8	12221	22	AAK71585	Human immune/haema
45	277	9.8	30393	22	AAK67239	Human immune/haema

ALIGNMENTS

RESULT 1  
AAC55312  
ID AAC55312 standard; cDNA; 2818 BP.  
XX AC AAC55312;  
XX DT 05-FEB-2001 (first entry)  
XX DE Human activation-induced cytidine deaminase encoding cDNA SEQ ID NO:7.  
XX DE Human activation-induced cytidine deaminase; AID: cytidine deaminase;  
KW Immune related disease; allergy; allergic disease; antiallergic;  
KW antianemic; antitachytic; ophthalmological; anti-Hiv; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IgA nephritis; gamma-globulinaemia; atopie dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ss.  
XX OS Homo sapiens.

XX FH Key Location/Qualifiers  
XX FT CDS 80..676  
XX FT /\*tag= a  
XX FT /product= "activation-induced cytidine deaminase"  
XX PN WO200058480-A1.  
XX XX 05-OCT-2000.

PF 28-MAR-2000; 2000WO-JP01918.  
 XX 29-MAR-1999; 99JP-0087192.  
 PR 24-JUN-1999; 99JP-0178999.  
 PR 27-DEC-1999; 99JP-0371382.  
 XX (NISB) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 XX Honjo T, Muramatsu M;  
 XX WPI; 2000-611715/58.  
 DR P-PSDB; AAB24198.  
 DR  
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 XX  
 PS Claim 3; Page 135-139; 174pp; Japanese.  
 CC The present sequence encodes human activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
 CC antinaemic, antisthmatic, ophthalmological, anti-HIV and  
 CC dermatological activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
 CC disease, DeGeorge disease, ataxia telangiectasia, common variable  
 CC immunodeficiency disorder, MHC (major histocompatibility class) class  
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
 CC encoding AID may be used for gene therapy and the antibodies to the AID  
 CC protein may be used for diagnosis and treatment of these disorders.  
 XX  
 SQ Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 other;  
 Query Match 100.0%; Score 2818; DB 21; Length 2818;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2818; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGAGAACCATTAATTAAGTGAGATTTTCTGGCCGTGAGACTTGCAGGGAGGCAAGA 60  
 DB 1 AGAGAACCATTAATTAAGTGAGATTTTCTGGCCGTGAGACTTGCAGGGAGGCAAGA 60  
 QY 61 AGACACTCTGGACACCATATGACAGCCCTTGTGATGAACCGGAGGAAGTTCTTTACCA 120  
 DB 61 AGACACTCTGGACACCATATGACAGCCCTTGTGATGAACCGGAGGAAGTTCTTTACCA 120  
 QY 121 ATTCAAAATATGTCGGCTAAGGCTGCGCTGAGACCTACCTGTCTACGTAGTGAA 180  
 DB 121 ATTCAAAATATGTCGGCTAAGGCTGCGCTGAGACCTACCTGTCTACGTAGTGAA 180  
 QY 181 GAGCGTGACAGTGCTACATCCCTTTTCACTGGACTTTGGTATCTTCGCAATGAAGCGG 240  
 DB 181 GAGCGTGACAGTGCTACATCCCTTTTCACTGGACTTTGGTATCTTCGCAATGAAGCGG 240  
 QY 241 CTGCCACGTGAATTCCTCTCCGCTACATCTCGGACTGGGACTAGACCCCTGGCCG 300  
 DB 241 CTGCCACGTGAATTCCTCTCCGCTACATCTCGGACTGGGACTAGACCCCTGGCCG 300  
 QY 301 CTGCTACCGCGTCACTGGTTACCTCTCGGAGCCCTGCTACGACTGTGCCCGGACATGT 360  
 DB 301 CTGCTACCGCGTCACTGGTTACCTCTCGGAGCCCTGCTACGACTGTGCCCGGACATGT 360  
 QY 361 GGCCGACTTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTA 420  
 DB 361 GGCCGACTTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTA 420  
 QY 421 CTTCTGTGAGACCGAAGGCTGAGCCCGGAGGGCTGCGGCGCTGCACCGCCCGGGGT 480  
 DB 421 CTTCTGTGAGACCGAAGGCTGAGCCCGGAGGGCTGCGGCGCTGCATCCCGCGGGGT 480

DB 421 CTTCTGTGAGACCGAAGGCTGAGCCCGGAGGGCTGCGGCGCTGCACCGCCCGGGGT 480  
 QY 481 GCAAAATAGCATCATGACCTTCAAGATTAATTTTACTGCTGCTGAATACTTTTCTAGAAAA 540  
 DB 481 GCAAAATAGCATCATGACCTTCAAGATTAATTTTACTGCTGCTGAATACTTTTCTAGAAAA 540  
 QY 541 CCATGAAAGAACTTTCAAGCCCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCTCCAG 600  
 DB 541 CCATGAAAGAACTTTCAAGCCCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCTCCAG 600  
 QY 601 ACAGCTTCGGCGCATCTTTTGGCCCTGTATGAGGTGATGAGCTTACGAGACCATTTTCG 660  
 DB 601 ACAGCTTCGGCGCATCTTTTGGCCCTGTATGAGGTGATGAGCTTACGAGACCATTTTCG 660  
 QY 661 TACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACCATGAAATATCTCTGCTG 720  
 DB 661 TACTTTGGGACTTTGATAGCAACTTCCAGGAATGTCACACACCATGAAATATCTCTGCTG 720  
 QY 721 AAGACAGTGGATAAAAAACAGTCTTCAAGTCTTCTCTGTTTATTTCTTCAACTCTCAC 780  
 DB 721 AAGACAGTGGATAAAAAACAGTCTTCAAGTCTTCTCTGTTTATTTCTTCAACTCTCAC 780  
 QY 781 TTTCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTTAAAAAGATCTATGCTTG 840  
 DB 781 TTTCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTTAAAAAGATCTATGCTTG 840  
 QY 841 AAAATAGAGAAGGAACACAGGCTCTGCCAGGAGCTGCTGCAATTTGGTGCAGTTTGAAT 900  
 DB 841 AAAATAGAGAAGGAACACAGGCTCTGCCAGGAGCTGCTGCAATTTGGTGCAGTTTGAAT 900  
 QY 901 GCAACATTTGCCCTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATCTTAAAGTGT 960  
 DB 901 GCAACATTTGCCCTACTGGGAATAACAGAACTGCAGGACCTGGGAGCATCTTAAAGTGT 960  
 QY 961 CAACGTTTCTATGACTTTTAGTGGATGAGAGCAGAAAGTAGATCTTAAAAAGCATG 1020  
 DB 961 CAACGTTTCTATGACTTTTAGTGGATGAGAGCAGAAAGTAGATCTTAAAAAGCATG 1020  
 QY 1021 GTGAGAGGATCAAAATGTTTATATCAACATCTTTTATTTATTTGATTTCATTTGAGTTAAC 1080  
 DB 1021 GTGAGAGGATCAAAATGTTTATATCAACATCTTTTATTTATTTGATTTCATTTGAGTTAAC 1080  
 QY 1081 AGTGGTGTAGTGATAGATTTTCTATCTTTTCCCTTGCAGCTTTACTTTCAAGTAAAC 1140  
 DB 1081 AGTGGTGTAGTGATAGATTTTCTATCTTTTCCCTTGCAGCTTTACTTTCAAGTAAAC 1140  
 QY 1141 AAACCTCTCCATCAGGCCATGATCTATAGGACCTCTCTAATGAGAGTATCTGGGTGATGT 1200  
 DB 1141 AAACCTCTCCATCAGGCCATGATCTATAGGACCTCTCTAATGAGAGTATCTGGGTGATGT 1200  
 QY 1201 GACCCCAAAACCATCTCTCCAAAGCATTAATATCAATCATGCGCTGTATGTTTAAATCAG 1260  
 DB 1201 GACCCCAAAACCATCTCTCCAAAGCATTAATATCAATCATGCGCTGTATGTTTAAATCAG 1260  
 QY 1261 CAGAACATGTTTTTATGTTGTACAAAAGAGATTGTTATGGGTGGGATGAGGATATA 1320  
 DB 1261 CAGAACATGTTTTTATGTTGTACAAAAGAGATTGTTATGGGTGGGATGAGGATATA 1320  
 QY 1321 GACCATGCTGGTGCACCTTCAAGCTACTTTTAAATAAAGAGTCTTAAATGGGAGGAGGAC 1380  
 DB 1321 GACCATGCTGGTGCACCTTCAAGCTACTTTTAAATAAAGAGTCTTAAATGGGAGGAGGAC 1380  
 QY 1381 TGTGAACAAGACACCCCTTAATATGGGTGTGATGCTGAAGTAGCAAAATCTTCTGGAACGC 1440  
 DB 1381 TGTGAACAAGACACCCCTTAATATGGGTGTGATGCTGAAGTAGCAAAATCTTCTGGAACGC 1440  
 QY 1441 AAACCTCTTTTAAAGAGTCCCTAAATTTAGAACACCCCAAACTTCCACATATCAATTA 1500  
 DB 1441 AAACCTCTTTTAAAGAGTCCCTAAATTTAGAACACCCCAAACTTCCACATATCAATTA 1500  
 QY 1501 GCAAAACAATTTGGAAGGAAGTTGCTTGAATCTTGGGAGAGGAGAAATCTATTGGCTCTCGT 1560  
 DB 1501 GCAAAACAATTTGGAAGGAAGTTGCTTGAATCTTGGGAGAGGAGAAATCTATTGGCTCTCGT 1560

[illegible]

Qy	2641	TTTATGACAGCAAAATTTGGCTCTGGCTCAGTTTCAATCAGTTTAAATGATAATAAT	2760
Db	2641	TTTATGACAGCAAAATTTGGCTCTGGCTCAGTTTCAATCAGTTTAAATGATAATAAT	2760
Qy	2701	TTTGGAGCGTGTGAAGATAAAATACCAAAATAAAAATATATAAAAAGTGATTTATATGAAGT	2760
Db	2701	TTTGGAGCGTGTGAAGATAAAATACCAAAATAAAAATATATAAAAAGTGATTTATATGAAGT	2760
Qy	2761	TAAATAAAAAAATCAGTATGATGAATAAACTTGAAAAAAATAAAAAAATAAAAAA	2818
Db	2761	TAAATAAAAAAATCAGTATGATGAATAAACTTGAAAAAAATAAAAAAATAAAAAA	2818
RESULT 2			
AAC55314			
ID	AAC55314	standard; DNA: 6564 BP.	
XX			
AC	AAC55314;		
XX			
DT	05-FEB-2001	(first entry)	
XX			
DE	Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:10.		
XX			
KW	Activation-induced cytidine deaminase; AID; cytidine deaminase;		
KW	immune related disease; allergy; allergic disease; antiallergic;		
KW	antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;		
KW	gene therapy; B cell associated immune system disorder; food allergy;		
KW	immunodeficiency disease; immunoglobulin A deficiency disease; asthma;		
KW	IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;		
KW	drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;		
KW	ataxia telangiectasia; common variable immunodeficiency disorder;		
KW	major histocompatibility class II deficiency disease;		
KW	auto immunodeficiency syndrome; IgG subclass selection disorder; ds.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200058480-A1.		
XX			
PD	05-OCT-2000.		
XX			
PF	28-MAR-2000; 2000WO-JP01918.		
XX			
PR	29-MAR-1999; 99JP-0087192.		
PR	24-JUN-1999; 99JP-0178999.		
PR	27-DEC-1999; 99JP-0371382.		
XX			
PA	(NLSB ) JAPAN TOBACCO INC.		
FA	(HONJ/) HONJO T.		
PI	Honjo T, Muramatsu M;		
XX			
DR	WPI; 2000-611715/58.		
XX			
PT	Nucleic acid encoding activation induced cytidine deaminase, useful as		
PT	a target for drug development for immune-related diseases including		
PT	allergies--		
PS	Claim 17; Page 145-150; 174pp; Japanese.		
XX			
CC	The present invention describes an activation-induced cytidine deaminase		
CC	(AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and		
CC	has cytidine activity similar to APOBEC-1. AID has antiallergic,		
CC	antianemic, antiasthmatic, ophthalmological, anti-HIV and		
CC	dermatological activities, and can be used in gene therapy. AID		
CC	polynucleotides are useful in methods for identifying drugs for the		
CC	treatment of B cell associated immune system disorders, immunodeficiency		
CC	diseases and allergies, such as immunoglobulin A (IgA) deficiency		
CC	disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic		
CC	colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen		
CC	disease, DiGeorge disease, ataxia telangiectasia, common variable		
CC	immunodeficiency disorder, MHC (major histocompatibility class) class		
CC	II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated		
CC	IgE disorder, and IgG subclass selection disorder. The DNA sequences		

CC	encoding AID may be used for gene therapy and the antibodies to the AID									
CC	protein may be used for diagnosis and treatment of these disorders. The									
CC	present sequence represents a genomic DNA sequence of human AID.									
XX										
SQ	Sequence 6564 BP; 1909 A; 1358 C; 1383 G; 1914 T; 0 other;									
	Query Match	77.2%	Score	2174.6;	DB	21;	Length	6564;		
	Best Local Similarity	99.8%	Pred.	No. 0;						
	Matches 2177;	Conservative	0;	Mismatches	4;	Indels	0;	Gaps	0;	
QY	617	CTTTTCCCTGCTATGAGTGGTATGACTTACGAGAGCCATTTTCGTACTTTGGGACTTTGA	676							
DB	3735	CTACAGCCCTGATGAGTGGTATGACTTACGAGAGCCATTTTCGTACTTTGGGACTTTGA	3794							
QY	677	TAGCAACTTCCAGGAATGTCACACAGATGAATATCTCTGCTGAAGACAGTGGATAAAA	736							
DB	3795	TAGCAACTTCCAGGAATGTCACACAGATGAATATCTCTGCTGAAGACAGTGGATAAAA	3854							
QY	737	AACAGTCCCTCAAGTCTCTCTGTTTATTTCTTCAACTCTCATTCTTCTTAGAGTTTACA	796							
DB	3855	AACAGTCCCTCAAGTCTCTCTGTTTATTTCTTCAACTCTCATTCTTCTTAGAGTTTACA	3914							
QY	797	GAATAAATATTTATATACGACTCTTTAAAGAGATCTATGCTTGAATAAGAGAAAC	856							
DB	3915	GAATAAATATTTATACGACTCTTTAAAGAGATCTATGCTTGAATAAGAGAAAC	3974							
QY	857	ACAGGTCTGGCCAGGACGTGCTGCAATTGGTGCAGTTTGAATGCAACATTCGCCCTA	916							
DB	3975	ACAGGTCTGGCCAGGACGTGCTGCAATTGGTGCAGTTTGAATGCAACATTCGCCCTA	4034							
QY	917	CTGGGAATACAGAACTGCAGGACCTGGGAGCATCTCTAAAGTGTCAACGTTTTCATGA	976							
DB	4035	CTGGGAATACAGAACTGCAGGACCTGGGAGCATCTCTAAAGTGTCAACGTTTTCATGA	4094							
QY	977	CTTTTAGTGGATGAGACAGAGTGTAGATCTCTAAAGATGAGTGGTGTAGTGATA	1036							
DB	4095	CTTTTAGTGGATGAGACAGAGTGTAGATCTCTAAAGATGAGTGGTGTAGTGATA	4154							
QY	1037	TTTTTATCAACATCCTTTATTTATTTGATTTCAATTTGAGTTAAACAGTGGTGTAGTGATA	1096							
DB	4155	TTTTTATCAACATCCTTTATTTATTTGATTTCAATTTGAGTTAAACAGTGGTGTAGTGATA	4214							
QY	1097	GATTTTTCATTTCTTTCCCTTGACGTTTACTTTTCAAGTAACACAACTCTTCCATCAGG	1156							
DB	4215	GATTTTTCATTTCTTTCCCTTGACGTTTACTTTTCAAGTAACACAACTCTTCCATCAGG	4274							
QY	1157	CCATGATCTATAGACCTCTCTAATGAGAGTATCTGGGTGATTTGACGTAACAGTCTC	1216							
DB	4275	CCATGATCTATAGACCTCTCTAATGAGAGTATCTGGGTGATTTGACGTAACAGTCTC	4334							
QY	1217	TCCAAAGCATTAATATCAATCATGCGCTGTATGTTTAAATCAGCAGAGCATGTTTAA	1276							
DB	4335	TCCAAAGCATTAATATCAATCATGCGCTGTATGTTTAAATCAGCAGAGCATGTTTAA	4394							
QY	1277	TGTTTGTACAAAGAGATGTTATGGTGGGATGGAGGTATAGACCATGCAATGTCAC	1336							
DB	4395	TGTTTGTACAAAGAGATGTTATGGTGGGATGGAGGTATAGACCATGCAATGTCAC	4454							
QY	1337	CTTCAAGCTACTTTAATAAGGATCTTAAATGGGAGGAGGACTCTGAACAGACACCC	1396							
DB	4455	CTTCAAGCTACTTTAATAAGGATCTTAAATGGGAGGAGGACTCTGAACAGACACCC	4514							
QY	1397	TAATAATGGTGTGATGCTGAAGTAGCAAACTCTTCTGAAACGCAAACTCTTTTAAGGAA	1456							
DB	4515	TAATAATGGTGTGATGCTGAAGTAGCAAACTCTTCTGAAACGCAAACTCTTTTAAGGAA	4574							
QY	1457	GTCCCTAATTTAGAAACACCCACAACTTCAACATATCATTAATAGCAAACTTTGAAGG	1516							
DB	4575	GTCCCTAATTTAGAAACACCCACAACTTCAACATATCATTAATAGCAAACTTTGAAGG	4634							
QY	1517	AAGTCTCTTGAATGTTGGGAGAGGAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA	1576							
DB	4635	AAGTCTCTTGAATGTTGGGAGAGGAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA	4694							

QY	1577	GAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGTGTCTCTCCCAAG	1636							
DB	4695	GAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGTGTCTCTCCCAAG	4754							
QY	1637	GTATATTAATATATAGAGAGTTGTGACAAAACAGAAATGATAAGCTGCGAACCGTGC	1696							
DB	4755	GTATATTAATATATAGAGAGTTGTGACAAAACAGAAATGATAAGCTGCGAACCGTGC	4814							
QY	1697	ACAGCTCATAGTCTTAGCTGCTTGGGAGGTTGAGGAGGAGGATGCTTTGAACACAGT	1756							
DB	4815	ACAGCTCATAGTCTTAGCTGCTTGGGAGGTTGAGGAGGAGGATGCTTTGAACACAGT	4874							
QY	1757	GTTCAAGGCCAGCTTGGGCAACATACAAAGATCCTGCTCTCAAAAAAATAAAAAA	1816							
DB	4875	GTTCAAGGCCAGCTTGGGCAACATACAAAGATCCTGCTCTCAAAAAAATAAAAAA	4934							
QY	1817	AAGAAAGAGAGAGGGCGGCGTGGTGGCTACGCCCTGTAAATCCACGACCTTTGGGAGC	1876							
DB	4935	AAGAAAGAGAGAGGGCGGCGTGGTGGCTACGCCCTGTAAATCCACGACCTTTGGGAGC	4994							
QY	1877	CGAGCGGCGGATCACCTGCTGCTCAGGAGTTTGAACACGAGCTTGGCCAAACATGGCAAA	1936							
DB	4995	CGAGCGGCGGATCACCTGCTGCTCAGGAGTTTGAACACGAGCTTGGCCAAACATGGCAAA	5054							
QY	1937	CCCGCTCTGTACTCAAAATGCAAAAAATTAGCCAGGCTGTAGCAGGACCTCTTAATCCC	1996							
DB	5055	CCCGCTCTGTACTCAAAATGCAAAAAATTAGCCAGGCTGTAGCAGGACCTCTTAATCCC	5114							
QY	1997	AGTACTTGGGAGGCTGAGGAGGAGAAATCGCTTTGAACCCAGGAGTGGAGTTGCAGTA	2056							
DB	5115	AGTACTTGGGAGGCTGAGGAGGAGAAATCGCTTTGAACCCAGGAGTGGAGTTGCAGTA	5174							
QY	2057	AGCTGAGATGCTGCGCTTGCACCTCCAGCCTGGCGGACAGAGCAAGACTCTGTCTCAGAA	2116							
DB	5175	AGCTGAGATGCTGCGCTTGCACCTCCAGCCTGGCGGACAGAGCAAGACTCTGTCTCAGAA	5234							
QY	2117	AAAAAATAAAG	2176							
DB	5235	AAAAAATAAAG	5294							
QY	2177	AAGCATTCGAGAGAAATTTGCTTTATCCAAACAAATGTAAGGAGGCAATTAAGGATCCC	2236							
DB	5295	AAGCATTCGAGAGAAATTTGCTTTATCCAAACAAATGTAAGGAGGCAATTAAGGATCCC	5354							
QY	2237	TATTTGCTCTTTTGGTGTCTATTGTCCTTAACAACAGTCTTTGACAGTGAGAAAAATA	2296							
DB	5355	TATTTGCTCTTTTGGTGTCTATTGTCCTTAACAACAGTCTTTGACAGTGAGAAAAATA	5414							
QY	2297	TTCAGAAATACCATATCCCTGTGCCGTTATTACCTAGCAACCCCTTGCATTAAGAGATGAGC	2356							
DB	5415	TTCAGAAATACCATATCCCTGTGCCGTTATTACCTAGCAACCCCTTGCATTAAGAGATGAGC	5474							
QY	2357	AGATCCACAGGAAACCTTGAATGCACAACTGCTTATTTTAATCTTATTTGTACATAGTT	2416							
DB	5475	AGATCCACAGGAAACCTTGAATGCACAACTGCTTATTTTAATCTTATTTGTACATAGTT	5534							
QY	2417	TGTAAGAGAGTTAAAAATTTGTTACTTCATGTTATTTATATATTTATTTTTCGCT	2476							
DB	5535	TGTAAGAGAGTTAAAAATTTGTTACTTCATGTTATTTATATATTTATTTTTCGCT	5594							
QY	2477	CTAATGATTTTATTAACATGATTTCCCTTTTCTGATATATTTGAAATGGAGTCTCAAGC	2536							
DB	5595	CTAATGATTTTATTAACATGATTTCCCTTTTCTGATATATTTGAAATGGAGTCTCAAGC	5654							
QY	2537	TTCAATAATTTTAACTTTAGAAATGATTCATAACACAGTATGTAATTTGTAACATTC	2596							
DB	5655	TTCAATAATTTTAACTTTAGAAATGATTCATAACACAGTATGTAATTTGTAACATTC	5714							
QY	2597	AGTAATGGTGTACGAGGACATTTCTTGTATTTTGTATTTTGTATTTTGTACACAGCAAT	2656							
DB	5715	AGTAATGGTGTACGAGGACATTTCTTGTATTTTGTATTTTGTATTTTGTACACAGCAAT	5774							

QY 2657 TCGTCTTGCTCACTTTCAATCAGTTAAATAAATGATAAATAATTTTGAAGCTGTGAAG 2716  
 Db 5775 TCGTCTTGCTCACTTTCAATCAGTTAAATAAATGATAAATAATTTTGAAGCTGTGAAG 5834

QY 2717 ATAAATACCAATTAATAATATAAAGTGAATTTATATGAAGTTAAATAAATAATCAG 2776  
 Db 5835 ATAAATACCAATTAATAATATAAAGTGAATTTATATGAAGTTAAATAAATAATCAG 5894

QY 2777 TATGATGAATAAAGTTGAAA 2797  
 Db 5895 TATGATGAATAAAGTTGAGA 5915

RESULT 3  
 ID AAC55339 standard; DNA; 11204 BP.  
 XX AAC55339;  
 AC AAC55339;  
 DT 05-FEB-2001 (first entry)  
 DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.  
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; antiallergic;  
 KW antianemic; antialsthmatic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease;  
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.  
 KW Homo sapiens.  
 OS Homo sapiens.  
 XX WO200058480-A1.  
 PN 05-OCT-2000.  
 PD 28-MAR-2000; 2000WO-JP01918.  
 PF 29-MAR-1999; 99JP-0087192.  
 PR 24-JUN-1999; 99JP-0178999.  
 PR 27-DEC-1999; 99JP-0371382.  
 XX (NIBS) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 PI Honjo T, Muramatsu M;  
 XX WPI; 2000-611715/58.  
 DR Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 XX Claim 17; Page 163-170; 174pp; Japanese.  
 PS The present invention describes an activation-induced cytidine deaminase  
 XX (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
 CC antianemic, antialsthmatic, ophthalmological, anti-HIV and  
 CC dermatological activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
 CC immunodeficiency disorder, MHC (major histocompatibility class) class  
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences

CC encoding AID may be used for gene therapy and the antibodies to the AID  
 CC protein may be used for diagnosis and treatment of these disorders. The  
 CC present sequence represents a genomic DNA sequence of human AID.  
 XX Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 other;  
 SQ Query Match 77.2%; Score 2174.6; DB 21; Length 11204;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 2177; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 617 CTTTGGCCCTCTATGAGTTGATGACTTTACAGAGCGCATTTCTGACTTTGGGACTTTGA 676  
 Db 8950 CTACAGCCCTCTATGAGTTGATGACTTTACAGAGCGCATTTCTGACTTTGGGACTTTGA 9009

QY 677 TAGCAACTTCCAGGAATGTCACACAGCATGAATATCTCTGCTGAAGACAGTGAATAAA 736  
 Db 9010 TAGCAACTTCCAGGAATGTCACACAGCATGAATATCTCTGCTGAAGACAGTGAATAAA 9069

QY 737 AACAGTCCCTTCAAGTCTTCTCTGTTTATTCTTCAACTCTCTCAGTTTCTTAGAGTTTACA 796  
 Db 9070 AACAGTCCCTTCAAGTCTTCTCTGTTTATTCTTCAACTCTCTCAGTTTCTTAGAGTTTACA 9129

QY 797 GAAAAATATTTATATAGACTCTTTTAAAGATCTATGCTTTGAAATAGAGAAGAAC 856  
 Db 9130 GAAAAATATTTATATAGACTCTTTTAAAGATCTATGCTTTGAAATAGAGAAGAAC 9189

QY 857 ACAGTCTGCGCCAGGAGCTGCTCAATTTGGTGCAGTTTGAATGCAACATTTGCCCTTA 916  
 Db 9190 ACAGTCTGCGCCAGGAGCTGCTCAATTTGGTGCAGTTTGAATGCAACATTTGCCCTTA 9249

QY 917 CTGGGAATAACAGAACTGCGAGACCTGGGAGCATCTTAAAGTGTCAACGTTTTCTTATGA 976  
 Db 9250 CTGGGAATAACAGAACTGCGAGACCTGGGAGCATCTTAAAGTGTCAACGTTTTCTTATGA 9309

QY 977 CTTTATAGTAGGATGAGACAGAGTAGATCCTTAAAGAGCATGGTGAGAGGATCAATG 1036  
 Db 9310 CTTTATAGTAGGATGAGACAGAGTAGATCCTTAAAGAGCATGGTGAGAGGATCAATG 9369

QY 1037 TTTTATATCAACATCCTTTTATTATTGATTCTTATTTGAGTTTAAACAGTGGTGTAGTGATA 1096  
 Db 9370 TTTTATATCAACATCCTTTTATTATTGATTCTTATTTGAGTTTAAACAGTGGTGTAGTGATA 9429

QY 1097 GATTTTCTTATTTTCCCTTGAGCTTTTCAAGTAAACACAACTCTTCCATCAGG 1156  
 Db 9430 GATTTTCTTATTTTCCCTTGAGCTTTTCAAGTAAACACAACTCTTCCATCAGG 9489

QY 1157 CCATGATCTATAGGACCTCCTTAATGAGAGTATCTGGGTGATTTGTGACCCCAACCATCTC 1216  
 Db 9490 CCATGATCTATAGGACCTCCTTAATGAGAGTATCTGGGTGATTTGTGACCCCAACCATCTC 9549

QY 1217 TCCAAAGCATTAATATCCCAATCATGCGTGTATGTTTAAATCAGCAGAGAGCATGTTTTTA 1276  
 Db 9550 TCCAAAGCATTAATATCCCAATCATGCGTGTATGTTTAAATCAGCAGAGAGCATGTTTTTA 9609

QY 1277 TGTTTGTACAAAAGAGATTGTTATGGGTGGGATGGAGGTATAGACCATGATGTCAC 1336  
 Db 9610 TCTTGTACAAAAGAGATTGTTATGGGTGGGATGGAGGTATAGACCATGATGTCAC 9669

QY 1337 CTTCAAGCTACCTTTTAAATAGGATCTTAAATGGGAGGAGGACTGTGAACAAGACACC 1396  
 Db 9670 CTTCAAGCTACCTTTTAAATAGGATCTTAAATGGGAGGAGGACTGTGAACAAGACACC 9729

QY 1397 TAATATGGGTGATGCTCTGAAGTAGCAAAATCTCTGGAACGCAAACTCTTTTAAAGGA 1456  
 Db 9730 TAATATGGGTGATGCTCTGAAGTAGCAAAATCTCTGGAACGCAAACTCTTTTAAAGGA 9789

QY 1457 GTCCCTTAATTTAGAACACCCCAAACTTCAATATCATATAATAGCAAACTTTGAAGG 1516  
 Db 9790 GTCCCTTAATTTAGAACACCCCAAACTTCAATATCATATAATAGCAAACTTTGAAGG 9849

QY 1517 AAGTTGCTTGAATGTTGGGAGGAGGAAATCTATTGGCTCTCTGGGTCTCTTTCATCTCA 1576  
 Db 9850 AAGTTGCTTGAATGTTGGGAGGAGGAAATCTATTGGCTCTCTGGGTCTCTTTCATCTCA 9909

Qy 1577 GAAATGCAATCAGGTCAAGGTTTGCTACATTTTGTATGCTGTGATGCTTCTCCCAAAG 1636  
Dy 9910 GAAATGCAATCAGGTCAAGGTTTGCTACATTTTGTATGCTGTGATGCTTCTCCCAAAG 9969  
Qy 1637 GTATATTAACTATATAAGAGAGTTGTGCACAAACAGAAATGATAAGCTGCGAACCGTGGC 1696  
Dy 9970 GTATATTAACTATATAAGAGAGTTGTGCACAAACAGAAATGATAAGCTGCGAACCGTGGC 10029  
Qy 1697 ACAGGCTCATAGTTCTAGCTGCTTGGAGGTTGAGGAGGAGGATGCTTGAACACAGGT 1756  
Dy 10030 ACAGGCTCATAGTTCTAGCTGCTTGGAGGTTGAGGAGGAGGATGCTTGAACACAGGT 10089  
Qy 1757 GTTCAAGCGCAGCTGGGCAACATACAAAGATCTGCTCTCAAAAAAATAAAAAA 1816  
Dy 10090 GTTCAAGCGCAGCTGGGCAACATACAAAGATCTGCTCTCAAAAAAATAAAAAA 10149  
Qy 1817 AAGAAAGAGAGAGCGCGGGGCTGGTGGCTCAGCCCTGTAAATCCACGACCTTTGGGAGGC 1876  
Dy 10150 AAGAAAGAGAGAGCGCGGGGCTGGTGGCTCAGCCCTGTAAATCCACGACCTTTGGGAGGC 10209  
Qy 1877 CGAGCCGCGGAGTCACTGCTGCTCAGGAGTTTTCAGACACGCTTGGCCACATGGCAAAA 1936  
Dy 10210 CGAGCCGCGGAGTCACTGCTGCTCAGGAGTTTTCAGACACGCTTGGCCACATGGCAAAA 10269  
Qy 1937 CCGCGTCTGTACTCAAAATGCAAAATTAGCCAGCGTGTAGCAGGCACTGTGAATCCC 1996  
Dy 10270 CCGCGTCTGTACTCAAAATGCAAAATTAGCCAGCGTGTAGCAGGCACTGTGAATCCC 10329  
Qy 1997 AGCTACTTGGAGGCTGAGCAGAGAGTAATCGTTGAACCCAGGAGTGGAGTTGTCAGTA 2056  
Dy 10330 AGCTACTTGGAGGCTGAGCAGAGAGTAATCGTTGAACCCAGGAGTGGAGTTGTCAGTA 10389  
Qy 2057 AGCTGAGATCGTCCGCTGTCACCTGCTGCGGCAAGCAAGCAAGTCTGCTCAGAA 2116  
Dy 10390 AGCTGAGATCGTCCGCTGTCACCTGCTGCGGCAAGCAAGCAAGTCTGCTCAGAA 10449  
Qy 2117 AAAAAAATAAAG 2176  
Dy 10450 AAAAAAATAAAG 10509  
Qy 2177 AAGCATTCGAAGGAATTTGCTTTTATCCAAACAAATGTAAAGAGCCCAATAAAGGATCCC 2236  
Dy 10510 AAGCATTCGAAGGAATTTGCTTTTATCCAAACAAATGTAAAGAGCCCAATAAAGGATCCC 10569  
Qy 2237 TATTGTCTCTTTTGGTGTCTATTGTGCTTAACTGCTTTGACAGTGAGAGAAATA 2296  
Dy 10570 TATTGTCTCTTTTGGTGTCTATTGTGCTTAACTGCTTTGACAGTGAGAGAAATA 10629  
Qy 2297 TTCAGATAACCATATCCCTGTGCTGCTTATTTACCTAGCAACCCCTTGCATTAAGAGATGAGC 2356  
Dy 10630 TTCAGATAACCATATCCCTGTGCTGCTTATTTACCTAGCAACCCCTTGCATTAAGAGATGAGC 10689  
Qy 2357 AGATCCACAGAAACTTGAATGCACACACTGCTTATTTTAACTTATTTGATACATAGTT 2416  
Dy 10690 AGATCCACAGAAACTTGAATGCACACACTGCTTATTTTAACTTATTTGATACATAGTT 10749  
Qy 2417 TGTAAAGAGTTAAATAATCTTACTCATGATTTTCAATTTATTTATTTATTTTTCGCT 2476  
Dy 10750 TGTAAAGAGTTAAATAATCTTACTCATGATTTTCAATTTATTTATTTATTTTTCGCT 10809  
Qy 2477 CTAATGATTTTATTAACATGATTTCCCTTCTGATATATTGAAATGGAGTCTCAAGC 2536  
Dy 10810 CTAATGATTTTATTAACATGATTTCCCTTCTGATATATTGAAATGGAGTCTCAAGC 10869  
Qy 2537 TTTCAATATTTTAACTTTAGAAATGATTTCTTAATACAGCTATGTAATTTGATACATGTC 2596  
Dy 10870 TTTCAATATTTTAACTTTAGAAATGATTTCTTAATACAGCTATGTAATTTGATACATGTC 10929  
Qy 2597 AGTAATGCTGCTACGAGGCAATTTCTTCTGATTTTAACTTTTATGTAACCTTTTATGACAGCAAT 2656  
Dy 10930 AGTAATGCTGCTACGAGGCAATTTCTTCTGATTTTAACTTTTATGTAACCTTTTATGACAGCAAT 10989

Qy 2657 TGCTTCTGGCTCACTTTCAATCACTTAAATAAATGATAAATAATTTTGGAGCTGTGAAG 2716  
Dy 10990 TGCTTCTGGCTCACTTTCAATCACTTAAATAAATGATAAATAATTTTGGAGCTGTGAAG 11049  
Qy 2717 ATAAATACCAAAATAAATAAATAAATAAAGTGATTTATATGAAGTTAAATAAATAAATCAG 2776  
Dy 11050 ATAAATACCAAAATAAATAAATAAATAAAGTGATTTATATGAAGTTAAATAAATAAATCAG 11109  
Qy 2777 TATGATGGAATAAATTTGAAA 2797  
Dy 11110 TATGATGGAATAAATTTGAGA 11130  
RESULT 4  
AAC55319  
ID AAC55319 standard; DNA: 2172 BP.  
AC AAC55319;  
XX  
XX 05-FEB-2001 (first entry)  
DT Human activation-induced cytidine deaminase exon 5 SEQ ID NO:15.  
DE  
XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; antiallergic;  
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; aschma;  
KW Iga nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.  
XX Homo sapiens.  
OS  
XX WO200058480-A1.  
PN  
XX  
XX 05-OCT-2000.  
PD  
XX  
XX 28-MAR-2000; 2000WO-JP01918.  
PF  
XX  
XX 29-MAR-1999; 99JP-0087192.  
PR  
XX 24-JUN-1999; 99JP-0178999.  
PR  
XX 27-DEC-1999; 99JP-0371382.  
PR  
XX (NISR) JAPAN TOBACCO INC.  
PA (HONJ)/ HONJO T.  
PA  
XX Honjo T, Muramatsu M;  
PI  
XX WPI; 2000-611715/58.  
DR  
XX  
XX Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
PT  
XX  
PS Claim 18; Page 152-153; 174pp; Japanese.  
XX  
XX The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
CC antianemic, antiasthmatic, ophthalmological, anti-HIV and  
CC dermatological activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (IGA) deficiency  
CC disease, Iga nephritis, gamma-globulinemia, atopic dermatitis, allergic  
CC colitis, aschma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class) class  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC Ige disorder, and Ige subclass selection disorder. The DNA sequences



Qy	623	CCCCGTGATGAGGTTGATGACTTACGAGACGCCATTTGCTACTTTTGGGACTTTTGATAGCAA	682
Db	1	CCCCGTGATGAGGTTGATGACTTACGAGACGCCATTTGCTACTTTTGGGACTTTTGATAGCAA	60
Qy	683	CTTCCAGGAATGTCACACACGATCAAAATATCTCTGCTGAACACAGTGTGATAAAAAACAGT	742
Db	61	CTTCCAGGAATGTCACACACGATCAAAATATCTCTGCTGAACACAGTGTGATAAAAAACAGT	120
Qy	743	CCTTCAAGTCTTCTCTGTTTTATTCTTCAACTCTCACTTTCTTTAGAGTTTACAGAAAA	802
Db	121	CCTTCAAGTCTTCTCTGTTTTATTCTTCAACTCTCACTTTCTTTAGAGTTTACAGAAAA	180
Qy	803	ATATTTATATACGACTCTTTAAAAGATCTATGTTCTTGAAATATAGAGAAACACAGGT	862
Db	181	ATATTTATATACGACTCTTTAAAAGATCTATGTTCTTGAAATATAGAGAAACACAGGT	240
Qy	863	CTGCCAGGAGCGTGTGCAATTCGGTCAGTTTTGAATGCAACATGTGCCCTACTGGA	922
Db	241	CTGCCAGGAGCGTGTGCAATTCGGTCAGTTTTGAATGCAACATGTGCCCTACTGGA	300
Qy	923	ATAACAGAACTGCAAGGACCTGGGAGCATCTTAAAGTGTCAACGTTTTTCTATGACTTTTA	982
Db	301	ATAACAGAACTGCAAGGACCTGGGAGCATCTTAAAGTGTCAACGTTTTTCTATGACTTTTA	360
Qy	983	GGTAGGATGAGAGCAGAGAAGGTAGATCCTAAAAGCATGGTGAGAGGATCAAAATGTTTTA	1042
Db	361	GGTAGGATGAGAGCAGAGAAGGTAGATCCTAAAAGCATGGTGAGAGGATCAAAATGTTTTA	420
Qy	1043	TATCAACATCCTTTATTTATTTGATTCATTTGAGTTTAAAGTGTGAGTTAGTGATAGATTTT	1102
Db	421	TATCAACATCCTTTATTTATTTGATTCATTTGAGTTTAAAGTGTGAGTTAGTGATAGATTTT	480
Qy	1103	TCTATTTCTTTTCCCTTGACGTTTACTTTCAAGTAAACAAACTCTTCCATCAGGCCATGA	1162
Db	481	TCTATTTCTTTTCCCTTGACGTTTACTTTCAAGTAAACAAACTCTTCCATCAGGCCATGA	540
Qy	1163	TCTATAGACCTCTTAATGAGAGTATCTGGGTGATGTGAGCCCAAACCATCTCTCCAAA	1222
Db	541	TCTATAGACCTCTTAATGAGAGTATCTGGGTGATGTGAGCCCAAACCATCTCTCCAAA	600
Qy	1223	GCATTAAATCCAAATATGCGCTGTATGTTTTTAATCAGCAAGCATGTTTTATGTTTG	1282
Db	601	GCATTAAATCCAAATATGCGCTGTATGTTTTTAATCAGCAAGCATGTTTTATGTTTG	660
Qy	1283	TACAAAAGAAGATTGTTATGGGTGGGATGGAGGTATAGACCATGCTGTCACCTCAA	1342
Db	661	TACAAAAGAAGATTGTTATGGGTGGGATGGAGGTATAGACCATGCTGTCACCTCAA	720
Qy	1343	GCTACTTTTATAAAGGATCTTAAATGGGCAGGAGGACTGTGAACAAGACACCCCTATAAA	1402
Db	721	GCTACTTTTATAAAGGATCTTAAATGGGCAGGAGGACTGTGTGAACAAGACACCCCTATAAA	780
Qy	1403	TGGGTGATGTCGAAGTACCAATCTTCTGGAACGCAACTCTTTTAGGAAGTCCCT	1462
Db	781	TGGGTGATGTCGAAGTACCAATCTTCTGGAACGCAACTCTTTTAGGAAGTCCCT	840
Qy	1463	AATTTAGAAACACCCCAAACTTCCACATATCATTAATAGCAAACTTTGGAAGGAGTTG	1522
Db	841	AATTTAGAAACACCCCAAACTTCCACATATCATTAATAGCAAACTTTGGAAGGAGTTG	900
Qy	1523	CTTGAATGTTGGGGAGAGGAAAATCTATTTGGCTCTCTGGGTCTCTTCACTCAGAAATG	1582
Db	901	CTTGAATGTTGGGGAGAGGAAAATCTATTTGGCTCTCTGGGTCTCTTCACTCAGAAATG	960

QY	2663	TGCTCAGTTTCAATCAGTTAAATAAATGATAAATAATTTTGAAGCTGTGAAGATAAAA	2722	PR	01-SEP-2000;	2000US-0229345.
Db	2041	TGGCTCAGTTTCAATCAGTTAAATAAATGATAAATAATTTTGAAGCTGTGAAGATAAAA	2100	PR	05-SEP-2000;	2000US-0229509.
QY	2723	TACCAATAAATAAATAAAGTATTTATGAACTTAAATAAATAAATAAATCAGTATGAT	2782	PR	06-SEP-2000;	2000US-0230437.
Db	2101	TACCAATAAATAAATAAAGTATTTATGAACTTAAATAAATAAATAAATCAGTATGAT	2160	PR	08-SEP-2000;	2000US-0230438.
QY	2783	GGAATAAACTTG 2794		PR	08-SEP-2000;	2000US-0231242.
Db	2161	GGAATAAACTTG 2172		PR	08-SEP-2000;	2000US-0231243.
				PR	08-SEP-2000;	2000US-0231244.
				PR	08-SEP-2000;	2000US-0231413.
				PR	08-SEP-2000;	2000US-0231414.
				PR	08-SEP-2000;	2000US-0232080.
				PR	08-SEP-2000;	2000US-0232081.
				PR	12-SEP-2000;	2000US-0231968.
				PR	14-SEP-2000;	2000US-0232397.
				PR	14-SEP-2000;	2000US-0232398.
				PR	14-SEP-2000;	2000US-0232399.
				PR	14-SEP-2000;	2000US-0232400.
				PR	14-SEP-2000;	2000US-0232401.
				PR	14-SEP-2000;	2000US-0233063.
				PR	14-SEP-2000;	2000US-0233064.
				PR	14-SEP-2000;	2000US-0233065.
				PR	21-SEP-2000;	2000US-0234223.
				PR	21-SEP-2000;	2000US-0234274.
				PR	25-SEP-2000;	2000US-0234997.
				PR	25-SEP-2000;	2000US-0234998.
				PR	26-SEP-2000;	2000US-0235484.
				PR	27-SEP-2000;	2000US-0235834.
				PR	27-SEP-2000;	2000US-0235836.
				PR	29-SEP-2000;	2000US-0236327.
				PR	29-SEP-2000;	2000US-0236367.
				PR	29-SEP-2000;	2000US-0236368.
				PR	29-SEP-2000;	2000US-0236369.
				PR	29-SEP-2000;	2000US-0236370.
				PR	02-OCT-2000;	2000US-0236802.
				PR	02-OCT-2000;	2000US-0237037.
				PR	02-OCT-2000;	2000US-0237038.
				PR	02-OCT-2000;	2000US-0237039.
				PR	02-OCT-2000;	2000US-0237040.
				PR	13-OCT-2000;	2000US-0239935.
				PR	13-OCT-2000;	2000US-0239937.
				PR	20-OCT-2000;	2000US-0240960.
				PR	20-OCT-2000;	2000US-0241221.
				PR	20-OCT-2000;	2000US-0241785.
				PR	20-OCT-2000;	2000US-0241786.
				PR	20-OCT-2000;	2000US-0241787.
				PR	20-OCT-2000;	2000US-0241808.
				PR	20-OCT-2000;	2000US-0241809.
				PR	20-OCT-2000;	2000US-0241826.
				PR	01-NOV-2000;	2000US-0244617.
				PR	08-NOV-2000;	2000US-0246474.
				PR	08-NOV-2000;	2000US-0246475.
				PR	08-NOV-2000;	2000US-0246476.
				PR	08-NOV-2000;	2000US-0246477.
				PR	08-NOV-2000;	2000US-0246478.
				PR	08-NOV-2000;	2000US-0246523.
				PR	08-NOV-2000;	2000US-0246524.
				PR	08-NOV-2000;	2000US-0246525.
				PR	08-NOV-2000;	2000US-0246526.
				PR	08-NOV-2000;	2000US-0246527.
				PR	08-NOV-2000;	2000US-0246528.
				PR	08-NOV-2000;	2000US-0246532.
				PR	08-NOV-2000;	2000US-0246609.
				PR	08-NOV-2000;	2000US-0246610.
				PR	08-NOV-2000;	2000US-0246611.
				PR	08-NOV-2000;	2000US-0246613.
				PR	17-NOV-2000;	2000US-0249207.
				PR	17-NOV-2000;	2000US-0249208.
				PR	17-NOV-2000;	2000US-0249209.
				PR	17-NOV-2000;	2000US-0249210.
				PR	17-NOV-2000;	2000US-0249211.
				PR	17-NOV-2000;	2000US-0249212.
				PR	17-NOV-2000;	2000US-0249213.
				PR	17-NOV-2000;	2000US-0249214.

AAK81088 standard; DNA; 1665 BP.

AAK81088;

07-NOV-2001 (first entry)

Human immune/haematopoietic antigen genomic sequence SEQ ID NO:35900.

Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

cytostatic; gene therapy; vaccine; metastasis; ds.

Homo sapiens.

W0200157182-A2.

09-AUG-2001.

17-JAN-2001; 2001WO-US01354.

31-JAN-2000; 2000US-0179065.

04-FEB-2000; 2000US-0180828.

24-FEB-2000; 2000US-0184664.

02-MAR-2000; 2000US-0186350.

16-MAR-2000; 2000US-0189874.

17-MAR-2000; 2000US-0190076.

18-APR-2000; 2000US-0198123.

19-MAY-2000; 2000US-0205515.

07-JUN-2000; 2000US-0209467.

28-JUN-2000; 2000US-0214886.

30-JUN-2000; 2000US-0215135.

07-JUL-2000; 2000US-0216647.

07-JUL-2000; 2000US-0216880.

11-JUL-2000; 2000US-0217487.

11-JUL-2000; 2000US-0217496.

14-JUL-2000; 2000US-0218290.

26-JUL-2000; 2000US-0220963.

26-JUL-2000; 2000US-0220964.

14-AUG-2000; 2000US-0224518.

14-AUG-2000; 2000US-0224519.

14-AUG-2000; 2000US-0225213.

14-AUG-2000; 2000US-0225266.

14-AUG-2000; 2000US-0225267.

14-AUG-2000; 2000US-0225268.

14-AUG-2000; 2000US-0225270.

14-AUG-2000; 2000US-0225447.

14-AUG-2000; 2000US-0225757.

14-AUG-2000; 2000US-0225758.

14-AUG-2000; 2000US-0225759.

18-AUG-2000; 2000US-0226279.

22-AUG-2000; 2000US-0226681.

22-AUG-2000; 2000US-0226868.

22-AUG-2000; 2000US-0227182.

30-AUG-2000; 2000US-0228924.

01-SEP-2000; 2000US-0229287.

01-SEP-2000; 2000US-0229343.

01-SEP-2000; 2000US-0229344.



PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-483426/52.  
DR  
XX  
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
XX  
PS Disclosure; SEQ ID NO 35900; 3071pp + Sequence Listing; English.  
XX  
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/haematopoietic-related diseases, especially  
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/haematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 1665 BP; 497 A; 343 C; 365 G; 460 T; 0 other;  
  
Query Match 52.8%; Score 1489.2; DB 22; Length 1665;  
Best Local Similarity 99.8%; Pred. No. 2.4e-238;  
Matches 1491; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
  
QY 617 CTTTTCGCCCTGTATGAGTTGATGACTTACGAGACGCATTTTCGTACTTTGGACTTTGA 676  
D 172 CTACAGCCCTGTATGAGTTGATGACTTACGAGACGCATTTTCGTACTTTGGACTTTGA 231  
QY 677 TAGCAACTCCAGGAATGTCACACAGCATGAATATCTCTGCTGAAGACAGTGGATAAAA 736  
D 232 TAGCAACTCCAGGAATGTCACACAGCATGAATATCTCTGCTGAAGACAGTGGATAAAA 291  
QY 737 AACAGTCCCTCAAGTCTCTCTGTTTATCTTCAACTCTCACATTTCTTAGAGTTTACA 796  
D 292 AACAGTCCCTCAAGTCTCTCTGTTTATCTTCAACTCTCACATTTCTTAGAGTTTACA 351

QY 797 GAAAAAATATTTATATACGACTCTTTTAAAAAGATCTATCTCTTTGAAAAATAGAGAAGAAC 856  
D 352 GAAAAAATATTTATATACGACTCTTTTAAAAAGATCTATCTCTTTGAAAAATAGAGAAGAAC 411  
QY 857 ACAGGTCTGGCCAGGACGCTGCTGCAATTTGGTGCAGTTTGAATGCAACATTTCTCCCTA 916  
D 412 ACAGGTCTGGCCAGGACGCTGCTGCAATTTGGTGCAGTTTGAATGCAACATTTCTCCCTA 471  
QY 917 CTGGGAATAACAGAACTGCAGGACCTGGGAGGATCCTAAAGTGTCTCAACGTTTCTTATGA 976  
D 472 CTGGGAATAACAGAACTGCAGGACCTGGGAGGATCCTAAAGTGTCTCAACGTTTCTTATGA 531  
QY 977 CTTTATAGTATGAGATGAGACAGAGGTAGATCCTAAAAAGCATGGTGAGAGGATCAAAATG 1036  
D 532 CTTTATAGTATGAGATGAGACAGAGGTAGATCCTAAAAAGCATGGTGAGAGGATCAAAATG 591  
QY 1037 TTTTATATCAACATCCTTTTATTTGATTTCATTTGAGTTTAACTTACAGTGTGTAGTGATA 1096  
D 592 TTTTATATCAACATCCTTTTATTTGATTTCATTTGAGTTTAACTTACAGTGTGTAGTGATA 651  
QY 1097 GATTTTCTTATTTCTTTTCCCTTGACGTTTACTTTTCAAGTAAACAACTCTTCCATCAGG 1156  
D 652 GATTTTCTTATTTCTTTTCCCTTGACGTTTACTTTTCAAGTAAACAACTCTTCCATCAGG 711  
QY 1157 CCATGATCTATAGGACCTCCTTAATGAGAGTATCTGGGTGATTGTGACCCCAACCATCTC 1216  
D 712 CCATGATCTATAGGACCTCCTTAATGAGAGTATCTGGGTGATTGTGACCCCAACCATCTC 771  
QY 1217 TCCAAAGCATTAATATCCAAATCATGCGCTGTATGTTTAAATCAGCAGAGATGTTTTTA 1276  
D 772 TCCAAAGCATTAATATCCAAATCATGCGCTGTATGTTTAAATCAGCAGAGATGTTTTTA 831  
QY 1277 TGTGTGTACAAAAGAAAGATTGTTATGGGTGGGATGGAGGTATAGACCATGCTGTCAC 1336  
D 832 TGTGTGTACAAAAGAAAGATTGTTATGGGTGGGATGGAGGTATAGACCATGCTGTCAC 891  
QY 1337 CTTCAAGCTACTTTAATAAAGGATCTTAAATGGGAGGAGGACTGTGAACAAAGACACC 1396  
D 892 CTTCAAGCTACTTTAATAAAGGATCTTAAATGGGAGGAGGACTGTGAACAAAGACACC 951  
QY 1397 TAATTAATGGGTTGATGCTCTGAAGTAGCAAAATCTTCTGGAACGCAAACTCTTTTAAGGA 1456  
D 952 TAATTAATGGGTTGATGCTCTGAAGTAGCAAAATCTTCTGGAACGCAAACTCTTTTAAGGA 1011  
QY 1457 GTCCCTTAATTTAGAAACACCCCAAACTTCAATATATCAATATAGCAAACTTTGGAAGG 1516  
D 1012 GTCCCTTAATTTAGAAACACCCCAAACTTCAATATATAGCAAACTTTGGAAGG 1071  
QY 1517 AAGTTGCTTGAATGTTGGGAGAGAAATCTATTGGCTCTCGTGGGTCTCTTTCATCTCA 1576  
D 1072 AAGTTGCTTGAATGTTGGGAGAGAAATCTATTGGCTCTCGTGGGTCTCTTTCATCTCA 1131  
QY 1577 GAAATGCCAATCAGGTCRAAGTTTGTCTACATTTGCTGATGCTGCTGCTCTCCCAAG 1636  
D 1132 GAAATGCCAATCAGGTCRAAGTTTGTCTACATTTGCTGATGCTGCTGCTCTCCCAAG 1191  
QY 1637 GTATATTAATATATAAGAGAGTTGTGACAAAACAGAAATGATAAAGCTGCGAACCTGGC 1696  
D 1192 GTATATTAATATATAAGAGAGTTGTGACAAAACAGAAATGATAAAGCTGCGAACCTGGC 1251  
QY 1697 ACAGCTCATAGTTCTAGTCTGTTGGAGGTTGAGAGGAGGATGGCTTGAACACAGGT 1756  
D 1252 ACAGCTCATAGTTCTAGTCTGTTGGAGGTTGAGAGGAGGATGGCTTGAACACAGGT 1311  
QY 1757 GTTCAAGGCCAGCTGGGCAACATAACAGATCCTGTCTCTCAAAAAAANAANAANA 1816  
D 1312 GTTCAAGGCCAGCTGGGCAACATAACAGATCCTGTCTCTCAAAAAAANAANAANA 1371  
QY 1817 AAGAAGAGAGAGGGCGGCGGTGGTGCCTCAGCCCTGTAAATCCAGCACTTTGGAGGC 1876  
D 1372 AAGAAGAGAGAGGGCGGCGGTGGTGCCTCAGCCCTGTAAATCCAGCACTTTGGAGGC 1431  
QY 1877 CGAGCGGCGGATACCTGTGTGTCAGGAGTTTGAGACCAGCCTGGCCAAACATGGCAAA 1936



QY	843	AATAGAGAAGAACACAGGCTCTGGCCAGGAGCTGCTCAATTTGGTCAGTCTTTGAATGC	902	PR	18-APR-2000;	2000US-0198123.
Db	844	AGGACCGCCAGCAAGTATGTAAC-----TGAGCTTGCTGTGC	882	PR	19-MAY-2000;	2000US-0205515.
QY	903	AACATGTCCCTACT-GGGAATACAGAACTGCAGACCTGGGAGCATCTAAAGTCTC	961	PR	07-JUN-2000;	2000US-0209467.
Db	883	AACATCGGCATCTACTGGGAAACAGCATAAAGTTCAGAGCTTTGGTCTGAATGATGCTC	942	PR	28-JUN-2000;	2000US-0214886.
QY	962	AACGTTTCTTATGACTTTAGGTAGGATGAGACGAGAGGTAGATCTCTAAAAGCATGG	1021	PR	30-JUN-2000;	2000US-0215135.
Db	943	TTTTTTTCAACAGCAT-----GGAAAAGCATAT	971	PR	07-JUL-2000;	2000US-0216647.
QY	1022	TGAGAGGATCAAAATGTTTTATATCAACATCCTTTATTTATTTGATTCATTTGAGTTAAACA	1081	PR	11-JUL-2000;	2000US-0217487.
Db	972	GGAGACGACACACAGATTTGTTACACCCACCTGTGTTCCCTTGATTCATTTGAATCTCA	1031	PR	14-JUL-2000;	2000US-0218290.
QY	1082	GTGGTGTAGTAGTAGATTTTCTATCTTTCCCTTGACGTTTACTTTTCAAGTAAACACA	1141	PR	26-JUL-2000;	2000US-0220963.
Db	1032	GGGTATCAGTACGAGGATCTTCTATTCTTCCCTCTAAGGCTCACTTTCAGG-----	1084	PR	26-JUL-2000;	2000US-0220963.
QY	1142	AACCTCTTCATCAGCCATGATCTATAGACCTCCTAATGAGATATCTGGGTGATTGTG	1201	PR	14-AUG-2000;	2000US-0220964.
Db	1085	---GGTCTTTTCTGACAAGGTCAGGGGCTGTCTACAGTCTCTGTCTGAGCAAT----	1137	PR	14-AUG-2000;	2000US-0224518.
QY	1202	ACCCCAAAACCATCTCTCCAAAGCATTAATATCCAAATCATGGCTGTATGTTTAAATCAGC	1261	PR	14-AUG-2000;	2000US-0224519.
Db	1138	-CACAAGCCATCTCTCAAAACATTAATACTACGGCACATGCTGTATGTTTCA-----	1191	PR	14-AUG-2000;	2000US-0225213.
QY	1262	AGAAGCATGTTTTATGTTTGTACAAAAGAGATGTTTATGGTGGGATGGAGGTATAG	1321	PR	14-AUG-2000;	2000US-0225214.
Db	1192	-----CTGTCGCTGCTTTTTCACATTTGTGTGAAGGGCTTGGGGTGGGATTGA	1245	PR	14-AUG-2000;	2000US-0225266.
QY	1322	ACCATGCATGGTCACCTTCAAGCTACTTTAATAAGGATCTTAAATGGGAGGAGACT	1381	PR	14-AUG-2000;	2000US-0225447.
Db	1246	AGAATGCAGATCGGCTCTGGTGTATTTCAATAAAGGATCTTAAATGCAGATGAGGACT	1305	PR	14-AUG-2000;	2000US-0225757.
QY	1382	GTGAACAGACACCCCTAATATGGTTGATGTCTGAAGTAGCAAAATCTTCTGGAACGCA	1441	PR	14-AUG-2000;	2000US-0225758.
Db	1306	ACGAGAATCACTCTGAATGATGTTTCAAGATGTTTCAAGGATCTTAAATGCAGATGAGGACT	1365	PR	14-AUG-2000;	2000US-0225759.
QY	1442	AACCTTTTT 1450		PR	14-AUG-2000;	2000US-0225759.
Db	1366	GACTCTTTT 1374		PR	14-AUG-2000;	2000US-0225759.
RESULT 7					14-AUG-2000;	2000US-0225759.
ID	AAK81089	standard; DNA; 574 BP.				2000US-0225759.
XX	AC	AAK81089;				2000US-0225759.
XX	AC	AAK81089;				2000US-0225759.
DT	07-NOV-2001	(first entry)				2000US-0225759.
DE	Human	immune/haematopoietic antigen genomic sequence SEQ ID NO:35901.				2000US-0225759.
XX	Human	immune; haematopoietic; immune/haematopoietic antigen; cancer;				2000US-0225759.
KW	cytostatic; gene therapy; vaccine; metastasis; ds.					2000US-0225759.
XX	Homo sapiens.					2000US-0225759.
XX	WO200157182-A2.					2000US-0225759.
PN	09-AUG-2001.					2000US-0225759.
PD	17-JAN-2001;	2001WO-US01354.				2000US-0225759.
PF	31-JAN-2000;	2000US-0179065.				2000US-0225759.
XX	04-FEB-2000;	2000US-0180628.				2000US-0225759.
XX	24-FEB-2000;	2000US-0184664.				2000US-0225759.
PR	02-MAR-2000;	2000US-0186350.				2000US-0225759.
PR	16-MAR-2000;	2000US-0189874.				2000US-0225759.
PR	17-MAR-2000;	2000US-0190076.				2000US-0225759.

PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249219.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249267.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251038.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;  
 XX WPI; 2001-483426/52.  
 XX

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -

XX Disclosure; SEQ ID NO 35901; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytosolic  
 CC activity, and can be used in gene therapy and vaccine production. (I)

CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/hematopoietic-related diseases, especially  
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 574 BP; 201 A; 80 C; 85 G; 208 T; 0 other;

Query Match 20.4%; Score 574; DB 22; Length 574;  
 Best Local Similarity 100.0%; Pred. No. 1.6e-86;  
 Matches 574; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2211 AATGTAAGGAGCCCAATAAGGGATCCCTATTGCTCTCTTTGGTGTCTATTGTCCTTAAC 2270  
 Db 1 AATGTAAGGAGCCCAATAAGGGATCCCTATTGCTCTCTTTGGTGTCTATTGTCCTTAAC 60  
 QY 2271 AACTGTCTTTGACAGTGAGAAAAATATTCAGATAACCATATCCCTGTGCCGTATTACC 2330  
 Db 61 AACTGTCTTTGACAGTGAGAAAAATATTCAGATAACCATATCCCTGTGCCGTATTACC 120  
 QY 2331 TAGCAACCCCTTGAATGAAGATGAGCAGATCCACAGAAACCTTGAATGCACAACTGTCT 2390  
 Db 121 TAGCAACCCCTTGAATGAAGATGAGCAGATCCACAGAAACCTTGAATGCACAACTGTCT 180  
 QY 2391 TATTTTAATCTTATTGTACATAAGTTTGTAAAGAGCTTAAATTTCTTACTTCATCTATT 2450  
 Db 181 TATTTTAATCTTATTGTACATAAGTTTGTAAAGAGCTTAAATTTCTTACTTCATCTATT 240  
 QY 2451 CATTTATATTTTATATTATTTTCGCTCTAATGATTTTATTAAACATGATTTCTTTTCT 2510  
 Db 241 CATTTATATTTTATATTATTTTCGCTCTAATGATTTTATTAAACATGATTTCTTTTCT 300  
 QY 2511 GATATATTGAATGGAGTCTCAAGCTTCATAAATTTATAAATTTAGAAATGATTTCTAAT 2570  
 Db 301 GATATATTGAATGGAGTCTCAAGCTTCATAAATTTATAAATTTAGAAATGATTTCTAAT 360  
 QY 2571 AACACGATGTAATTTGTAACATTTGCAGTAATGGTCTACGAGCCATTTCTTCTCATTT 2630  
 Db 361 AACACGATGTAATTTGTAACATTTGCAGTAATGGTCTACGAGCCATTTCTTCTCATTT 420  
 QY 2631 TTAGTAAACTTTTATGACAGCAAAATTTGCTTCTGCTCCTCACTTTCAATCAGTTAAAT 2690  
 Db 421 TTAGTAAACTTTTATGACAGCAAAATTTGCTTCTGCTCCTCACTTTCAATCAGTTAAAT 480  
 QY 2691 GATAAATAATTTTGGAGCTGTGAAGATAAATACCAATAAATAAATAAATAAATGATTT 2750  
 Db 481 GATAAATAATTTTGGAGCTGTGAAGATAAATACCAATAAATAAATAAATAAATGATTT 540  
 QY 2751 TATATGAAGTAAAAATAAAAAATCAGTATGATGG 2784  
 Db 541 TATATGAAGTAAAAATAAAAAATCAGTATGATGG 574

RESULT 8  
 AAK61819  
 ID AAK61819 standard; cDNA; 429 BP.  
 XX  
 AC AAK61819;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX  
 DE Human immune/hematopoietic antigen encoding cDNA SEQ ID NO:6879.  
 XX  
 KW Human; immune; hematopoietic; immune/hematopoietic antigen; cancer;

KW	cytostatic; gene therapy; vaccine; metastasis; ss.	PR	25-SEP-2000;	2000US-0234998.
XX	Homo sapiens.	PR	26-SEP-2000;	2000US-0235484.
OS		PR	27-SEP-2000;	2000US-0235834.
XX		PR	27-SEP-2000;	2000US-0235836.
PN	WO200157182-A2.	PR	29-SEP-2000;	2000US-0236327.
XX		PR	29-SEP-2000;	2000US-0236367.
PD	09-AUG-2001.	PR	29-SEP-2000;	2000US-0236368.
XX		PR	29-SEP-2000;	2000US-0236369.
PF	17-JAN-2001; 2001WO-US01354.	PR	29-SEP-2000;	2000US-0236370.
XX		PR	02-OCT-2000;	2000US-0236802.
PR	31-JAN-2000; 2000US-0179065.	PR	02-OCT-2000;	2000US-0237037.
PR	04-FEB-2000; 2000US-0180628.	PR	02-OCT-2000;	2000US-0237038.
PR	24-FEB-2000; 2000US-0184664.	PR	02-OCT-2000;	2000US-0237039.
PR	02-MAR-2000; 2000US-0186350.	PR	02-OCT-2000;	2000US-0237040.
PR	16-MAR-2000; 2000US-0189874.	PR	13-OCT-2000;	2000US-0239935.
PR	17-MAR-2000; 2000US-0190076.	PR	13-OCT-2000;	2000US-0239937.
PR	18-APR-2000; 2000US-0198123.	PR	20-OCT-2000;	2000US-0240960.
PR	19-MAY-2000; 2000US-0205515.	PR	20-OCT-2000;	2000US-0241221.
PR	07-JUN-2000; 2000US-0209467.	PR	20-OCT-2000;	2000US-0241785.
PR	28-JUN-2000; 2000US-0214886.	PR	20-OCT-2000;	2000US-0241786.
PR	30-JUN-2000; 2000US-0215135.	PR	20-OCT-2000;	2000US-0241787.
PR	07-JUL-2000; 2000US-0216647.	PR	20-OCT-2000;	2000US-0241808.
PR	07-JUL-2000; 2000US-0216880.	PR	20-OCT-2000;	2000US-0241809.
PR	11-JUL-2000; 2000US-0217487.	PR	20-OCT-2000;	2000US-0241826.
PR	11-JUL-2000; 2000US-0217496.	PR	01-NOV-2000;	2000US-0244617.
PR	14-JUL-2000; 2000US-0218290.	PR	08-NOV-2000;	2000US-0246474.
PR	26-JUL-2000; 2000US-0220963.	PR	08-NOV-2000;	2000US-0246475.
PR	26-JUL-2000; 2000US-0220964.	PR	08-NOV-2000;	2000US-0246476.
PR	14-AUG-2000; 2000US-0224518.	PR	08-NOV-2000;	2000US-0246477.
PR	14-AUG-2000; 2000US-0224519.	PR	08-NOV-2000;	2000US-0246478.
PR	14-AUG-2000; 2000US-0225213.	PR	08-NOV-2000;	2000US-0246523.
PR	14-AUG-2000; 2000US-0225214.	PR	08-NOV-2000;	2000US-0246524.
PR	14-AUG-2000; 2000US-0225266.	PR	08-NOV-2000;	2000US-0246525.
PR	14-AUG-2000; 2000US-0225267.	PR	08-NOV-2000;	2000US-0246526.
PR	14-AUG-2000; 2000US-0225268.	PR	08-NOV-2000;	2000US-0246527.
PR	14-AUG-2000; 2000US-0225270.	PR	08-NOV-2000;	2000US-0246528.
PR	14-AUG-2000; 2000US-0225447.	PR	08-NOV-2000;	2000US-0246532.
PR	14-AUG-2000; 2000US-0225757.	PR	08-NOV-2000;	2000US-0246609.
PR	14-AUG-2000; 2000US-0225758.	PR	08-NOV-2000;	2000US-0246610.
PR	14-AUG-2000; 2000US-0225759.	PR	08-NOV-2000;	2000US-0246611.
PR	18-AUG-2000; 2000US-0226279.	PR	08-NOV-2000;	2000US-0246613.
PR	22-AUG-2000; 2000US-0226681.	PR	17-NOV-2000;	2000US-0249207.
PR	22-AUG-2000; 2000US-0226868.	PR	17-NOV-2000;	2000US-0249208.
PR	22-AUG-2000; 2000US-0227182.	PR	17-NOV-2000;	2000US-0249209.
PR	23-AUG-2000; 2000US-0227009.	PR	17-NOV-2000;	2000US-0249210.
PR	30-AUG-2000; 2000US-0228924.	PR	17-NOV-2000;	2000US-0249211.
PR	01-SEP-2000; 2000US-0229287.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000; 2000US-0229343.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000; 2000US-0229344.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000; 2000US-0229345.	PR	17-NOV-2000;	2000US-0249215.
PR	05-SEP-2000; 2000US-0229509.	PR	17-NOV-2000;	2000US-0249216.
PR	05-SEP-2000; 2000US-0229513.	PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000; 2000US-0230437.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000; 2000US-0230438.	PR	17-NOV-2000;	2000US-0249244.
PR	08-SEP-2000; 2000US-0231242.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000; 2000US-0231243.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000; 2000US-0231244.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000; 2000US-0231413.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000; 2000US-0231414.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000; 2000US-0232080.	PR	17-NOV-2000;	2000US-0249300.
PR	08-SEP-2000; 2000US-0232081.	PR	01-DEC-2000;	2000US-0250160.
PR	12-SEP-2000; 2000US-0231968.	PR	01-DEC-2000;	2000US-0250391.
PR	14-SEP-2000; 2000US-0232397.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000; 2000US-0232398.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000; 2000US-0232399.	PR	05-DEC-2000;	2000US-0256719.
PR	14-SEP-2000; 2000US-0232400.	PR	06-DEC-2000;	2000US-0251479.
PR	14-SEP-2000; 2000US-0232401.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000; 2000US-0233063.	PR	08-DEC-2000;	2000US-0251868.
PR	14-SEP-2000; 2000US-0233064.	PR	08-DEC-2000;	2000US-0251869.
PR	14-SEP-2000; 2000US-0233065.	PR	08-DEC-2000;	2000US-0251989.
PR	21-SEP-2000; 2000US-0234223.	PR	08-DEC-2000;	2000US-0251990.
PR	21-SEP-2000; 2000US-0234274.	PR	11-DEC-2000;	2000US-0254097.
PR	25-SEP-2000; 2000US-0234997.	PR	05-JAN-2001;	2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX WPI: 2001-483426/52.  
DR P-PSDB: AAM89038.  
XX

XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
XX useful for preventing, diagnosing and/or treating cancers and  
XX metastasis.

XX Claim 1: SEQ ID NO 6879; 3071pp + Sequence Listing; English.

XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patient's own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting  
XX the nucleic acids into a host cell and culturing the cell to express the  
XX protein. (I) proteins and polynucleotides may be used to prevent,  
XX diagnose and treat immune/hematopoietic-related diseases, especially  
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
XX to AAK87694 represent human immune/hematopoietic antigen genomic  
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
XX represent sequences used in the exemplification of the present invention.

XX SQ Sequence 429 BP; 144 A; 79 C; 97 G; 109 T; 0 other;

Query Match 15.2%; Score 429; DB 22; Length 429;  
Best Local Similarity 100.0%; Pred. No. 1.8e-62;  
Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1386 ACAAGACACCCTAATGTTGATGCTCTGAAGTAGCAATCTCTGGAACGCAACT 1445  
DB 1 ACAAGACACCCTAATGTTGATGCTCTGAAGTAGCAATCTCTGGAACGCAACT 60  
QY 1446 CTTTAAAGGAAGTCCCTAATTTAGAACACCCCAACTTCACATATCATATTAAGCAAA 1505  
DB 61 CTTTAAAGGAAGTCCCTAATTTAGAACACCCCAACTTCACATATCATATTAAGCAAA 120  
QY 1506 CAATTGGAAGGAAGTTCCTGTAATGTTGGGAGGAGGAATCTATTGGCTCTCGTGGGTC 1565  
DB 121 CAATTGGAAGGAAGTTCCTGTAATGTTGGGAGGAGGAATCTATTGGCTCTCGTGGGTC 180  
QY 1566 TCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTTGTGATGC 1625  
DB 181 TCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTTGTGATGC 240  
QY 1626 TTCTCCCAAGGTATATTAAGTATTAAGAGAGTTGTGACAAACAGAAATGATAAAGCTG 1685  
DB 241 TTCTCCCAAGGTATATTAAGTATTAAGAGAGTTGTGACAAACAGAAATGATAAAGCTG 300  
QY 1686 CGAACCGTGCACACGCTCATAGTTCTAGCTGCTGGAGGTTGAGGAGGAGGATGGCT 1745  
DB 301 CGAACCGTGCACACGCTCATAGTTCTAGCTGCTGGAGGTTGAGGAGGAGGATGGCT 360  
QY 1746 TGAACACAGGTGTTCAAGGCCAGCTGGGCAACATACAGATCCTCTCTCAAAAAAAA 1805  
DB 361 TGAACACAGGTGTTCAAGGCCAGCTGGGCAACATACAGATCCTCTCTCAAAAAAAA 420  
QY 1806 AAAAAAAA 1814  
DB 421 AAAAAAAA 429

RESULT 9  
ABA08208

ID XX ABA08208 standard; DNA; 13862 BP.  
AC XX ABA08208;  
XX  
DT 11-JAN-2002 (first entry)  
DE XX  
DE XX  
XX Human ovarian and breast cancer associated polynucleotide SEQ ID NO 1003.  
XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;  
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antilucer;  
KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;  
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;  
KW neurological disease; infection; human; secreted protein; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200155325-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01345.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.



Qy	2026	CGCTTGAACCCAGGAGGTGGAGGTTGCAGTAAGCTGAGATCGTCCCGCTTGCACTCCAGCC	2085
Db	1868	CGTTTGAAGCCAGGAGCGGAGTTTGCACCGAGCCGAGATTGTGCCACTGCACCTCCAGCC	1927
Qy	2086	TGGCGCAGACAGCACAACACTCTCTCTCAGAAAAAAGAAAAAGAGAGAGAGAA	2145
Db	1928	TGGCGGATAA-AGCGAGACTCTGTCTCAGAAAAAAGAAAAAGAAAAAGAAAAA	1986
Qy	2146	AGACAACAATATTGGCAGAGAGAGGATGGGGAAGCATTTGCCAAGGAAATTGTG	2197
Db	1987	AGAAAAAATTAGCCAGCGGTGGTGCATGCACCTGTAGTCTAGCTACTTGGG	2038
RESULT 10			
AAL02789			
ID	AAL02789 standard; DNA; 13862 BP.		
AC	AAL02789;		
XX	21-NOV-2001 (first entry)		
XX	Human reproductive system related antigen DNA SEQ ID NO: 5477.		
DE	Human; reproductive system related antigen; reproductive system disorder;		
KW	cancer; gene therapy; ds.		
KW	Homo sapiens.		
OS	WO200155320-A2.		
XX	02-AUG-2001.		
XX	17-JAN-2001; 2001WO-US01339.		
XX	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	11-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	14-JUL-2000; 2000US-0217496.		
PR	26-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225447.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	18-AUG-2000; 2000US-0225759.		
PR	22-AUG-2000; 2000US-0226681.		
PR	22-AUG-2000; 2000US-0226868.		
PR	23-AUG-2000; 2000US-0227182.		
PR	30-AUG-2000; 2000US-0227009.		
PR	01-SEP-2000; 2000US-0228824.		
PR	17-NOV-2000; 2000US-0229287.		
PR	01-SEP-2000; 2000US-0229343.		
PR	01-SEP-2000; 2000US-0229344.		
PR	01-SEP-2000; 2000US-0229345.		
PR	05-SEP-2000; 2000US-0229509.		
PR	05-SEP-2000; 2000US-0229511.		
PR	06-SEP-2000; 2000US-0230437.		
PR	06-SEP-2000; 2000US-0230438.		
PR	08-SEP-2000; 2000US-0231242.		
PR	08-SEP-2000; 2000US-0231243.		
PR	08-SEP-2000; 2000US-0231244.		
PR	08-SEP-2000; 2000US-0231413.		
PR	08-SEP-2000; 2000US-0231414.		





1667	AAACAGAAATGATTAAGCTGCGAACCCTGGGCACACGCTCATAGTTCCTAGCTGCTTTGGGAGG	Qy	1726
1668			
1508	AAACTTAAATAAAGCTGGTGGCTCTGTGGCTCATGCTGTGATCCACGACACTTTGGGAGG	Db	1567
1727	TTCAGGAGGAGGAGTGGCTTGAACACAGGTGTTTCAAGGCAGGCTGGGCAACATACAAAG	Qy	1786
1568	CTGAGGCAGGAGGATCACTTTGAGGTCAAGAGTTTGAGACCAAGCGCTGGCCAAACATGGCGAA	Db	1627
1787	ATCCTGTCTCTCAAAAAAAAAAAAAAAAAAAAAAGAAAGA-CAGAGGGCGGGCGTGGTGGC	Qy	1845
1628	ACCCATCTCTACTATAAATACAAAAATTAGCCAGCATGGCGGGGCATGTGGTGGTGGC	Db	1687
1846	TCACGCGCTGTAATCCCAAGCACTTTGGGAGCGCCAGCGCGGATCACCTGTGTGTCAGGA	Qy	1905
1688	TCATGGCTGTAAATCCCAAGCACTTTGGGAGCGCCAGGTGGCGGATCACCTGAGGTCAAGGA	Db	1747
1906	GTTTGTAGACAGCCTGGCCAACTATGCAAAACCCCGTCTGTCTACTCAAAATGCAAAAATTA	Qy	1965
1748	GTTCAAGACAGCAGCTGGACAACATGGCAAAACCCCTGTCTCTATTAAAAATACAAAAATTA	Db	1807
1966	GCCAGGCGGTGATAGCAGGCACCTGTAAATCCCAAGCTACTTTGGGAGGCTGAGGCAGGAGAAT	Qy	2025
1808	GCCAGGCTTTATGGTAGGGCGCCTTAATCCCAAGCTACTCGGAAGGCTGAGGCAGGAGAAT	Db	1867
2026	CGCTTGAACCCAGGAGGTGGAGGTTGCAGTAAGCTGAGATCTGTGGCTTGCACCTCCAGCC	Qy	2085

```
Db      1868  CGTTTGAAGCAGGAGCGGAGTTTGCACCGAGCGGAGATTGTGCCACTGCCACTCCAGCC 1927
Qy      2086  TGGCGCAAGACAGCAAGACTCTGTCTCAGAAAAAAGAAAAAGAGAGAGAGAGAGAA 2145
Db      1928  TGGCGGATAA-AGCGAGACTCTGTCTCAGAAAAAAGAAAAAGAAAAAGAAAAA 1986
Qy      2146  AGAACAATATTTGGGAGAGAAGGATGGGGAAGCATTTGCCAAGGAAATTGTG 2197
Db      1987  AGAAAAATTAGCCAGCGTGTGGCAGTGCACCTGTAGTCCTAGTACTTGGG 2038

RESULT 12
ID ABA15724
XX AC ABA15724;
XX DT 23-JAN-2002 (first entry)
XX DE Human nervous system related polynucleotide SEQ ID NO 8055.
XX KW Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW immunosuppressive; antinflammatory; anti-HIV; antibacterial; vulnerary;
KW antiparkinsonian; antiskilling; antianaemic; antiarthritic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW antiallergic; antidiabetic; antilulcer; anticonvulsant; antifungal;
KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX OS Homo sapiens.
XX WO200159063-A2.
PN PD 16-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01334.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
```













CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.

CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 25012 BP; 6932 A; 4835 C; 5560 G; 7685 T; 0 other;

Query Match 10.3%; Score 291; DB 22; Length 25012;

Best Local Similarity 79.6%; Pred. No. 1.9e-39;

Matches 356; Conservative 0; Mismatches 90; Indels 1; Gaps 1;

QY 1681 AGCTCGAACCGTGGCACACGCTCATAGTTCTAGCTGCTGGGAGGTTGAGAGGGAGGA 1740

Db ||||| | | ||||| || | | | | | ||||| || ||||| || ||||| || |||||

19050 AGCTGGGAGGGTGGCTCATGCTCTAATCCCAACACTGTGGGAGGCCAAGCGGGAGGA 19109

QY 1741 TGGCTTGRACACAGTGTTCAGGCCAGCGCTGGGCAACATAACAAGATCCTGCTCT-CA 1799

Db || | | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

19110 TTGCTGAAGTCCAGAGTTCAAGACCGCTGGGAGCATAGCCAGACCCCGTCTCTATT 19169

QY 1800 AAAAAAAAAAAAAAAAAAAGAGAGAGCGCGGCTGGTGTCTACGCTGTAATC 1859

Db || | ||||| ||||| || | ||||| ||||| ||||| ||||| ||||| ||||| |||||

19170 TTTTAATTAATAAAAAAAAAAATGTAGGCCCGGCGACGGTGGCTCACGCTGTAATC 19229

QY 1860 CCAGCACTTTGGGAGCGCGCGGCGGATCACCCTGTGTGTCAGGAGTTTGAGACGAGCC 1919

Db ||||| ||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||

19230 CTAGCACTTTGGGAGGCTGAGCGGGTGGATCACCGAGTTCAGGAGTTCAAGACGAGCC 19289

QY 1920 TGGCCAAACATGGCAAAACCCGCTCTGTACTCAAAATGCAAAATTAGCCAGCGCTGGTAG 1979

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

19290 TGGCCAAACATGGTGAACCCCTGTCTCTACTAAAAATACAAAAATTAGCTGGCGGTGGTG 19349

QY 1980 CAGGCACCTGTAAATCCAGCTACTTTGGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGG 2039

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

19350 CAGGCGCTGTAAATTCAGCTACTTTGGGAGGCTGAGGAGGAGAAATAGCTTGAACCCAGG 19409

QY 2040 AGGTGGAGGTTGCAGTAAGCTGAGATCGTGCCGTTGCACTCCAGCCTGGGCGCAAGAGC 2099

Db || | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

19410 AGGCAGAGGTTATGTTGAGCCGAGATTTGCCATTGCACTCCAGCCTGGGCAACAAGAGC 19469

QY 2100 AAGACTCTGCTCAGAAAAAAAAAAAA 2126

Db || | | ||||| ||||| || ||||| || ||||| || ||||| || ||||| || |||||

19470 GAAACGCCATCTCCAAAAAAGATA 19496

Search completed: June 19, 2003, 00:37:09  
Job time : 580.918 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 23:34:44 ; Search time 3799.83 Seconds  
(without alignments)  
12010.756 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 2818  
Sequence: 1 agagaccatcatgaattga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:\*\*

- 1: em\_estba:\*\*
- 2: em\_esthbm:\*\*
- 3: em\_estlin:\*\*
- 4: em\_estlmu:\*\*
- 5: em\_estlov:\*\*
- 6: em\_estlpl:\*\*
- 7: em\_estro:\*\*
- 8: em\_hlc:\*\*
- 9: gb\_estli:\*\*
- 10: gb\_est2:\*\*
- 11: gb\_hlc:\*\*
- 12: gb\_est3:\*\*
- 13: gb\_est4:\*\*
- 14: gb\_est5:\*\*
- 15: em\_estfun:\*\*
- 16: em\_estom:\*\*
- 17: gb\_gss:\*\*
- 18: em\_gss\_hum:\*\*
- 19: em\_gss\_inv:\*\*
- 20: em\_gss\_pln:\*\*
- 21: em\_gss\_vrt:\*\*
- 22: em\_gss\_fun:\*\*
- 23: em\_gss\_mam:\*\*
- 24: em\_gss\_mus:\*\*
- 25: em\_gss\_other:\*\*
- 26: em\_gss\_pro:\*\*
- 27: em\_gss\_rod:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	850.6	30.2	856	9	AL559877
2	847	30.1	872	12	BG758510
3	809.4	28.7	953	14	BQ065440
4	799.2	28.4	1052	14	BQ055935
5	739.4	26.2	743	12	BG686133
6	705.6	25.0	820	12	BG757089

7	705.2	25.0	820	12	BG755526
8	681	24.2	693	12	BF975096
9	667.4	23.7	942	12	BF975166
10	663	23.5	664	12	BG754140
11	655.4	23.3	693	12	BG757392
12	627.6	22.3	670	12	BG341546
13	626	22.2	956	12	BF664355
c 14	572.6	20.3	604	10	AW978582
15	568	20.2	928	12	BG398364
16	562	19.9	570	10	AW504807
17	551.4	19.6	843	12	BG758815
18	546.8	19.4	1034	12	BG755005
19	540	19.2	541	12	BF238155
20	493.6	17.5	511	12	BG170756
21	471	16.7	517	12	BG170824
c 22	470.2	16.7	507	9	AL581406
c 23	467.8	16.6	528	9	AA954956
24	467.6	16.6	939	12	BG341819
25	456.4	16.2	889	12	BG686876
26	453.8	16.1	1140	12	BF664352
c 27	446.2	15.8	476	10	AW135547
c 28	442	15.7	442	9	AT016902
c 29	441.4	15.7	464	10	AW452648
30	411.4	14.6	757	9	AJ446140
31	409.8	14.5	696	9	AJ453647
c 32	406.2	14.4	422	10	AW263139
33	405.4	14.4	522	12	BG144705
34	369.4	13.1	374	10	AW401901
35	366.2	13.0	729	9	AJ450296
36	365.6	13.0	688	9	AJ450317
c 37	358.2	12.7	363	9	AI249134
38	353.4	12.5	413	10	AW768364
c 39	353	12.5	353	9	AA831307
40	352	12.5	650	9	AJ449745
c 41	330.4	11.7	332	9	AA504653
c 42	321	11.4	321	9	AI468242
c 43	317.8	11.3	364	10	AW182255
c 44	308.6	11.0	653	10	BB637360
45	291.8	10.4	623	9	AJ450295

ALIGNMENTS

RESULT 1	AL559877	AL559877	856 bp	mRNA	linear	EST 16-FEB-2001
LOCUS	AL559877	LTI_FLO11_BCl	Homo sapiens	CDNA clone	CS0DG003YB14	5 prime
DEFINITION	AL559877	, mRNA sequence.				
ACCESSION	AL559877					
VERSION	AL559877.1	GI:12905793				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	Li W.B., Gruber C., Jessee J. and Polayes D.					
TITLE	1 (bases 1 to 856)					
JOURNAL	Full-length cDNA libraries and normalization					
COMMENT	Unpublished (2001)					
CONTACT	Genoscope					
Genoscope	- Centre National de Sequencage					
BP 191	91006 EVRY cedex - France					
Email:	seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.					
FEATURES	Location/Qualifiers					
source	1..856					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="CS0DG003YB14"					
	/clone_lib="LTI_FLO11_BCl"					
	/sex="male"					
	/tissue_type="B cells from Burkitt lymphoma"					
	/lab_host="DH10B"					

/note="Vector: pCMVSPORT 6: 1st strand cDNA was primed with a NotI-oligo(dn) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 209 a 217 c 202 g 226 t 2 others  
ORIGIN

Query Match 30.2%; Score 850.6; DB 9; Length 856;  
Best Local Similarity 99.6%; Pred. No. 2.3e-90;  
Matches 850; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 62 GACACTCGGACACCACTATGGACAGCCTCTTGATGAACCGGAGAAAGTTCTTTACCAA 121  
Db 1 GACACTCGGACACCACTATGGACAGCCTCTTGATGAACCGGAGAAAGTTCTTTACCAA 60

Qy 122 TTCAAAATGTCGCTGGCTAGGGTTCAGGCTGGGCTGAGACCTACCTGCTAGTGAAG 181  
Db 61 TTCAAAATGTCGCTGGCTAGGGTTCAGGCTGGGCTGAGACCTACCTGCTAGTGAAG 120

Qy 182 AGGCGTGACAGTGCTACATCTTTTACACTGACCTTTGTTATCTTCCCAATAAGAACGGC 241  
Db 121 AGGCGTGACAGTGCTACATCTTTTACACTGACCTTTGTTATCTTCCCAATAAGAACGGC 180

Qy 242 TGGCAGCTGGAATGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCTTGGCGGC 301  
Db 181 TGGCAGCTGGAATGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCTTGGCGGC 240

Qy 302 TGCTACCGCTGACCTGGTTCACCTCTCGAGCCCTGCTGACGACTGTCGCGACATGTG 361  
Db 241 TGCTACCGCTGACCTGGTTCACCTCTCGAGCCCTGCTGACGACTGTCGCGACATGTG 300

Qy 362 GCCGACTTCTCGAGGGAACCCCACTCAGTCTGAGGATCTCACCGCGCGGCTCTAC 421  
Db 301 GCCGACTTCTCGAGGGAACCCCACTCAGTCTGAGGATCTCACCGCGCGGCTCTAC 360

Qy 422 TTCTGTGAGGACCGCAAGGCTGAGCCGCGGCTGCGGCGCTGACCGCGCGGCGGTG 481  
Db 361 TTCTGTGAGGACCGCAAGGCTGAGCCGCGGCTGCGGCGCTGACCGCGCGGCGGTG 420

Qy 482 CAAATAGCCATATGACCTTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAAC 541  
Db 421 CAAATAGCCATATGACCTTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAAC 480

Qy 542 CATGAAGAAGCTTTCAAAGCTGGAAGGCTGCATGAAATTCAGTCTGCTCTCCAGA 601  
Db 481 CATGAAGAAGCTTTCAAAGCTGGAAGGCTGCATGAAATTCAGTCTGCTCTCCAGA 540

Qy 602 CAGCTTCGGCGCATCTTTTGGCCCTGTATGAGGTTGATGACTTACGACGCAATTCGT 661  
Db 541 CAGCTTCGGCGCATCTTTTGGCCCTGTATGAGGTTGATGACTTACGACGCAATTCGT 600

Qy 662 ACTTTGGGACTTTGATAGCACTTCCAGGAATGTCACACGATGAAATATCTCTGCTGA 721  
Db 601 ACTTTGGGACTTTGATAGCACTTCCAGGAATGTCACACGATGAAATATCTCTGCTGA 660

Qy 722 AGACAGTGAATAAAACAGTCTCTCAAGTCTTCTGTTTATTTTCTCAACTCTCACT 781  
Db 661 AGACAGTGAATAAAACAGTCTCTCAAGTCTTCTGTTTATTTTCTCAACTCTCACT 720

Qy 782 TTCCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTGA 841  
Db 721 TTCCTTAGAGTTTACAGAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTGA 780

Qy 842 AAATAGAGAGNACACAGGCTGCGCCAGGAGCTGCTGCAATTTGGTGAGTTTGAATG 901  
Db 781 AAATAGAGAGNACACAGGCTGCGCCAGGAGCTGCTGCAATTTGGTGAGTTTGAATG 840

Qy 902 CAACATTTGCCCG 914

|||||  
841 CAACATTTGCCCG 853

RESULT 2  
BG758510 872 bp mRNA linear EST 15-MAY-2001  
LOCUS 602712721F1 NIH\_MGC\_48 Homo sapiens CDNA clone IMAGE:4853069 5',  
DEFINITION mRNA sequence.  
ACCESSION BG758510  
VERSION BG758510.1 GI:14069163  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 872)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cga@rs-research.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LCM1698 row: 1 column: 06  
High quality sequence stop: 836.  
Location/Qualifiers  
1..872  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4853069"  
/clone\_lib="NIH\_MGC\_48"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; CDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-CDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: This is a NIH\_MGC Library."  
BASE COUNT 211 a 221 c 212 g 228 t  
ORIGIN

Query Match 30.1%; Score 847; DB 12; Length 872;  
Best Local Similarity 99.3%; Pred. No. 5.9e-90;  
Matches 861; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 16 ATTGAAGTGAAGATTTTCTGGCCTGAGACTTGCAGGAGGCAAGAAGACACTCTGGACAC 75  
Db 2 ATTGAAGTGAAGATTTTCTGGCCTGAGACTTGCAGGAGGCAAGAAGACACTCTGGACAC 61

Qy 76 CACTATGGACAGCCTCTTGATGAACCGGAGGAGATTTCTTTACCAATTTCAAAATGTCCG 135  
Db 62 CACTATGGACAGCCTCTTGATGAACCGGAGGAGATTTCTTTACCAATTTCAAAATGTCCG 121

Qy 136 CTGGGCTAAGGGTCGGCGTGAGACCTACCTGTGCTAGTGAAGAGGCGTGCACAGTGC 195  
Db 122 CTGGGCTAAGGGTCGGCGTGAGACCTACCTGTGCTAGTGAAGAGGCGTGCACAGTGC 181

Qy 196 TACATCCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGACGCGTGCACGTGGAAT 255  
Db 182 TACATCCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGACGCGTGCACGTGGAAT 241

Qy 256 GCTCTCTCCCGCTACATCTCGGACTGGACCTTAGACCCCTGGCCGCTGTACCGCGTAC 315  
Db 182 TACATCCTTTTCACTGGACTTTTGGTTATCTTCGCAATAAGACGCGTGCACGTGGAAT 241

```
Db 242 GCTCTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGCCGCTGCTACCGGTCAC 301
QY 316 CTGGTTACCTCCTCGAGCCCTGCTAGCACTGTGCCGACATGTGGCGACTTCTGCG 375
Db 302 CTGGTTACCTCCTCGAGCCCTGCTAGCACTGTGCCGACATGTGGCGACTTCTGCG 361
QY 376 AGGGAACCCCAACCTCAGTCTGAGGACTCTTCCACCGCGGCTCTACTTCTGTGAGACCG 435
Db 362 AGGGAACCCCAACCTCAGTCTGAGGACTCTTCCACCGCGGCTCTACTTCTGTGAGACCG 421
QY 436 CAAGGCTGAGCCCGAGGGCTGCGGGCTGACCGCGCGGGTGCAAAATAGACCATCAT 495
Db 422 CAAGGCTGAGCCCGAGGGCTGCGGGCTGACCGCGCGGGTGCAAAATAGACCATCAT 481
QY 496 GACCTTCAAGATATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTT 555
Db 482 GACCTTCAAGATATATTTTACTGCTGGATACTTTTGTAGAAAACCATGAAGAATTT 541
QY 556 CAAGGCTGGGAGGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTTCGGCGCAT 615
Db 542 CAAGGCTGGGAGGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTTCGGCGCAT 601
QY 616 CTTTTTGGCCCTGTATGAGGTGATGACTTACGAGACGCATTTCTGACTTTGGGACTTTG 675
Db 602 CTTTTTGGCCCTGTATGAGGTGATGACTTACGAGACGCATTTCTGACTTTGGGACTTTG 661
QY 676 ATAGCAACTTCCAGGAATGTACACACAGATGAAATATCTCTGCTGAAGACAGTGGATAA 735
Db 662 ATAGCAACTTCCAGGAATGTACACACAGATGAAATATCTCTGCTGAAGACAGTGGATAA 721
QY 736 AAACAGTCTTCAAGTCTTCTCTGTTTATTTTCACTCTCACTTCTTCAAGTTTAC 795
Db 722 AAACAGTCTTCAAGTCTTCTCTGTTTATTTTCACTCTCACTTCTTCAAGTTTAC 781
QY 796 AGAAAAATATTTATATAGCACTCTTAAAGATCTATGCTTGAAGATAGAGAGAA 855
Db 782 AGAAAAATATTTATATAGCACTCTTAAAGATCTATGCTTGAAGATAGAGAGAA 841
QY 856 CACAGGCTGCGCAGGAGGCTGCTGCA 882
Db 842 CCCCCTGCTGG-CAGGAGGCTGCTGCA 867

RESULT 3
LOCUS BQ065440
DEFINITION AGENCOURT_6855061 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5929977
5', mRNA sequence.
ACCESSION BQ065440
VERSION BQ065440.1 GI:19894486
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 953)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/;
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue procurement: Lou Staudt
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC2108 row: P column: 10
High quality sequence scop: 634.
Location/Qualifiers
1..953
/organism="Homo sapiens"
```

```
BASE COUNT 238 a 236 c 233 g 246 t
ORIGIN
Query Match 28.7%; Score 809.4; DB 14; Length 953;
Best Local Similarity 98.5%; Pred. No. 1.3e-85;
Matches 828; Conservative 0; Mismatches 11; Indels 2; Gaps 1;

QY 17 TTGAAGTGAAGATTTTCTGGCCTGAGACTTGCAGGAGGCAAGACACACTCTGGACACC 76
Db 1 TTGAAGTGAAGATTTTCTGGCCTGAGACTTGCAGGAGGCAAGACACACTCTGGACACC 60
QY 77 ACTATGGACAGCCTCTTGATGAACCGAGGAAAGTTTCTTACCAATTCAAAAATGTCCGC 136
Db 61 ACTATGGACAGCCTCTTGATGAACCGAGGAAAGTTTCTTACCAATTCAAAAATGTCCGC 120
QY 137 TGGGCTAAGGGTGGCGGTGAGACCTACCTGTGCTAGTAGTGAAGAGGGGTGACAGTGT 196
Db 121 TGGGCTAAGGGTGGCGGTGAGACCTACCTGTGCTAGTAGTGAAGAGGGGTGACAGTGT 180
QY 197 ACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGATTG 256
Db 181 ACATCCTTTTCACTGGACTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGATTG 240
QY 257 CTCCTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCCGCTCACC 316
Db 241 CTCCTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCCGCTCACC 300
QY 317 TGGTTCACTCCTCGAGCCCCCTGCTACGACTGTGCCGACATGTGCCGCGACTTTCGCGA 376
Db 301 TGGTTCACTCCTCGAGCCCCCTGCTACGACTGTGCCGACATGTGCCGCGACTTTCGCGA 360
QY 377 GGAACCCCAACCTCAGTCTGAGGATCTTACCAGCGGCTCTACTTCTGCTGAGGACCGC 436
Db 361 GGAACCCCAACCTCAGTCTGAGGATCTTACCAGCGGCTCTACTTCTGCTGAGGACCGC 420
QY 437 AAGGCTGAGCCGAGGGGCTGCGGCGCTGCACCGCGCGGGTGC AAAATAGCCATCATG 496
Db 421 AAGGCTGAGCCGAGGGGCTGCGGCGCTGCACCGCGCGGGTGC AAAATAGCCATCATG 480
QY 497 ACCTTCAAAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTC 556
Db 481 ACCTTCAAAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTC 540
QY 557 AAAGCCTGGGAGGGCTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCCGGCGCATC 616
Db 541 AAAGCCTGGGAGGGCTGCATGAAAATTCAGTTCGTCTCTCCAGACAGCTTCCGGCGCATC 600
QY 617 CTTTTCGCCCTGTATGAGGTGTGATGACTTACGAGACGCATTTCTGTTGGGACTTTGA 676
Db 601 CTTTTCGCCCTGTATGAGGTGTGATGACTTACGAGACGCATTTCTGTTGGGACTTTGA 660
QY 677 TAGCACTTCCAGGAATGTACACACAGATGAATATCTCTGCTGAAGACAGTGGATAAAA 736
Db 661 TAGCACTTCCAGGAATGTACACACAGATGAATATCTCTGCTGAAGACAGTGGATAAAA 720
QY 737 ACAGCTCTCAAGTCTTCTGTTTTTATTTCACTCTCACTTCTTAGAGTTTACA 796
Db 721 AACAGTCTTCAAGTCTTCTGTTTTTATTTCTCACTCTCACTTCTTAGAGTTTAC 780
```

```
/db_xref="taxon:9606"
/clone IMAGE:5929977"
/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
```



```
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4766234"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT      176 a   197 c   188 g   182 t
ORIGIN

Query Match      26.2%; Score 739.4; DB 12; Length 743;
Best Local Similarity 99.9%; Pred. No. 2.1e-77;
Matches 740; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY  4 GAACCATCATTAATGAAGTGAAGATTTTCTGGCCTGAGACTTGCAGGGAGGCAAGA 63
Db  1 | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
2 GAACCATCATTAATGAAGTGAAGATTTTCTGGCCTGAGACTTGCAGGGAGGCAAGA 61
QY  64 CACTCTGGACACCATATGACAGCCTCTTGATGAACCGAGGAGGAGCTTTCTTACCAATT 123
Db  62 CACTCTGGACACCATATGACAGCCTCTTGATGAACCGAGGAGGAGCTTTCTTACCAATT 121
QY  124 CAAAATGTCCGTGGGCTAAGGCTGCGGTGAGACCTACCTGTGCTAGCTAGTGAAGAG 183
Db  122 CAAAATGTCCGTGGGCTAAGGCTGCGGTGAGACCTACCTGTGCTAGTGAAGAG 181
QY  184 GCGTGACAGTGTACATCTTTCTCACTGGACTTTGTTATCTTCGCAATAAAGACGGCTG 243
Db  182 GCGTGACAGTGTACATCTTTCTCACTGGACTTTGTTATCTTCGCAATAAAGACGGCTG 241
QY  244 CCACGTGGAATGTCTTCTCCGCTACATCTCGGACTGGGACTGAGACCTGAGCCTGGCGGCTG 303
Db  242 CCACGTGGAATGTCTTCTCCGCTACATCTCGGACTGGGACTGAGACCTGAGCCTGGCGGCTG 301
QY  304 CTACCGGCTACCTGGTTCACCTCTCGGACCCCTGCTACGACTGTGCCCGACATGTGGC 363
Db  302 CTACCGGCTACCTGGTTCACCTCTCGGACCCCTGCTACGACTGTGCCCGACATGTGGC 361
QY  364 CGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGGGGGCCTCTACTT 423
Db  362 CGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGGGGGCCTCTACTT 421
QY  424 CTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGCTGACCGCGCGGGGTGCA 483
Db  422 CTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGCTGACCGCGCGGGGTGCA 481
QY  484 AATAGCCATCATGACCTTCAAGATTAATTTTACTGCTGGAATACTTTTGTAGAAAACCA 543
Db  482 AATAGCCATCATGACCTTCAAGATTAATTTTACTGCTGGAATACTTTTGTAGAAAACCA 541
QY  544 TGAAGAATCTTCAAGCCCTGGGAGGCTGCATGAAATTCAGTTGCTCTCCAGACA 603
Db  542 TGAAGAATCTTCAAGCCCTGGGAGGCTGCATGAAATTCAGTTGCTCTCCAGACA 601
QY  604 GCTTCGGCGCATCTCTTTGCCCTGTATGAGGTTGATGACTTACGAGAGCGATTTTCGTAC 663
Db  602 GCTTCGGCGCATCTCTTTGCCCTGTATGAGGTTGATGACTTACGAGAGCGATTTTCGTAC 661
QY  664 TTTGGGACTTTGATAGCACTTCCAGGAATGTACACAGCATGAAATATCTTCGTGAAG 723
Db  662 TTTGGGACTTTGATAGCACTTCCAGGAATGTACACAGCATGAAATATCTTCGTGAAG 721
QY  724 ACAGTGGATAAAAAACAGTCC 744
Db  722 ACAGTGGATAAAAAACAGTCC 742
```

```
RESULT 6
LOCUS      BG757089
DEFINITION 602715124F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4855517 5',
mRNA sequence.
ACCESSION  BG757089
VERSION     BG757089.1 GI:14067742
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 820)
            NIH-MGC http://mgs.nci.nih.gov/.
AUTHORS     National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE        Unpublished (1999)
JOURNAL      Contact: Robert Strausberg, Ph.D.
COMMENT      Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCMI704 row: o column: 06
            High quality sequence stop: 675.

FEATURES             source
                    1..820
                        Location/Qualifiers
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:4855517"
                        /clone_lib="NIH_MGC_48"
                        /tissue_type="primary B-cells from tonsils (cell line)"
                        /lab_host="DH10B (phage-resistant)"
                        /notes="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
                        Site:2: EcoRI; cDNA made by oligo-dT priming.
                        Directionally cloned into EcoRI/XhoI sites using the
                        following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                        for average insert size 1.8kb. Library constructed by Ling
                        Hong in the laboratory of Gerald M. Rubin (University of
                        California, Berkeley) using ZAP-cDNA synthesis kit
                        (Stratagene) and Superscript II RT (Life Technologies).
                        Note: this is a NIH_MGC Library."

BASE COUNT      210 a   230 c   200 g   180 t
ORIGIN

Query Match      25.0%; Score 705.6; DB 12; Length 820;
Best Local Similarity 96.5%; Pred. No. 1.7e-73;
Matches 753; Conservative 0; Mismatches 24; Indels 3; Gaps 3;

QY  1 AGAGAACCATCATTAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGCAGGAGGCAAGA 60
Db  6 AGAGAACCATCATTAATTGAAGTGAGATTTTTCTGGCCTGAGACTTGCAGGAGGCAAGA 65
QY  61 AGACACTCTGGACACCATATGACAGCCTCTTGATGAACCGGAGGAGTTCTTTACCA 120
Db  66 AGACACTCTGGACACCATATGACAGCCTCTTGATGAACCGGAGGAGTTCTTTACCA 125
QY  121 ATTCAAAAATGTCCGCTGGGCTAAGGCTGCGCTGAGACCTTACCTGTGCTAGTAGTAA 180
Db  126 ATTCAAAAATGTCCGCTGGGCTAAGGCTGCGCTGAGACCTTACCTGTGCTAGTAGTAA 185
QY  181 GAGCGGTGACAGTGTACATCTTTTTCACATGGAGCTTTGGTGTATCTTCGCAATAGAACGG 240
Db  186 GAGCGGTGACAGTGTACATCTTTTTCACATGGAGCTTTGGTGTATCTTCGCAATAGAACGG 245
QY  241 CTGCCAGCTGGAATGTCTCTCTCCGCTACATCTCGGACTGCGACCTAGACCTCGGCCG 300
Db  246 CTGCCAGCTGGAATGTCTCTCTCCGCTACATCTCGGACTGCGACCTAGACCTCGGCCG 305
```

```
QY 301 CTGCTACCGCGGTCACTGGTTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGT 360
Db 306 CTGCTACCGCGGTCACTGGTTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGT 365
QY 361 GCGCGACTTTCTCGGAGGGAACCCCAACCTCACTCTGAGATCTTCAACCGCGGCTCTTA 420
Db 366 GCGCGACTTTCTCGGAGGGAACCCCAACCTCACTCTGAGATCTTCAACCGCGGCTCTTA 425
QY 421 CTTCTGTGAGACCGCAAGGCTGAGCGCGAGGGCTGCGGCGCTGCACCGCGCGGGT 480
Db 426 CTTCTGTGAGACCGCAAGGCTGAGCGCGAGGGCTGCGGCGCTGCACCGCGCGGGT 485
QY 481 GCAATAGCCATCATGACCTTCAAGATATATTTTACTGCTGGAATCTTTGTAGAAAA 540
Db 486 GCAATAGCCATCATGACCTTCAAGATATATTTTACTGCTGGAATCTTTGTAGAAAA 545
QY 541 CCATGAAGAACTTTCAAGCCCTGGAAGGCTGCATGAAATTCAGTTCTCTCCAG 600
Db 546 CCATGAAGAACTTTCAAGCCCTGGAAGGCTGCATGAAATTCAGTTCTCTCCAG 605
QY 601 ACAGCTTGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCATTTCG 660
Db 606 ACAGCTTGGCGCATCTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGCATTCG 664
QY 661 TACTTTGGGACTTTGTATAGCACTTCCAGGAATGTCACACA-CGATGAAATATCTCTGCT 719
Db 665 TACTTTAGGACTTTGTATAGCACTTCCAGGAATGTCACACAGGATGAAATATCTCTGCT 724
QY 720 GAAGACAGTGGATAA-AAACACGCTCTCAAGCTTCTCTGTTTATTTCTCAACTCTC 778
Db 725 GAAGACAGGCGATAACAAACGACCTCAAAGTCAACACTGAAGACAGCGCGTCAACGCC 784
```

## RESULT 7

BG755526

## LOCUS

602716206F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4856568 5', mRNA sequence.

## ACCESSION

BG755526

## VERSION

EST.

## KEYWORDS

human.

## SOURCE

Homo sapiens

## ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 820)

## AUTHORS

NIH-MGC http://mgi.nci.nih.gov/

## TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

http://image.llnl.gov

Plate: LCM1707 row: k column: 01

High quality sequence stop: 638.

## FEATURES

Location/Qualifiers

1..820

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4856568"

/clone\_lib="NIH\_MGC\_48"

/lab\_host="PH10B (phage-resistant)"

/tissue\_type="primary B-cells from tonsils (cell line)"

/notes="Organ: B-cells; Vector: pORF7; Site 1: XhoI;

Site 2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected &gt;500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library.

BASE COUNT 245 a 161 c 168 g 246 t  
ORIGIN

Query Match 25.0%; Score 705.2; DB 12; Length 820;  
Best Local Similarity 94.5%; Pred. No. 1.9e-73;  
Matches 775; Conservative 0; Mismatches 38; Indels 7; Gaps 4;

```
QY 634 GGTGTGACTTACGAGACGATTTTCGTACTTTGGGACTTTGTAGCAACTTCCAGGAAT 693
Db 2 GGTGTGACTTACGAGACGATTTTCGTACTTTGGGACTTTGTAGCAACTTCCAGGAAT 60
QY 694 GTCACACAGCATGAATATCTCTGCTGAACACAGTGGATAAAAACAGTCTCTCAAGTCT 753
Db 61 GTCACACAGCATGAATATCTCTGCTGAACACAGTGGATAAAAACAGTCTCTCAAGTCT 120
QY 754 TCTCTGTTTTTATTCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAAAATATTATATA 813
Db 121 TCTCTGTTTTTATTCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAAAATATTATATA 180
QY 814 CGACTCTTTAAAAAGATCTATGCTTTGAAAAATAGAGAAGAACACACAGGTCTGGCCAGGA 873
Db 181 CGACTCTTTAAAAAGATCTATGCTTTGAAAAATAGAGAAGAACACACAGGTCTGGCCAGGA 240
QY 874 CGTCTGCAATTCGTGCGAGTTTGAATGCAACATGTCGCCCTACTGGGAATAACAGAACT 933
Db 241 CGTCTGCAATTCGTGCGAGTTTGAATGCAACATGTCGCCCTACTGGGAATAACAGAACT 300
QY 934 GCAGGACCTGGGAGCATCTTAAAGTGTCAACAGTTTTTCTATGACTTTTAGTAGGATGAG 993
Db 301 GCAGGACCTGGGAGCATCTTAAAGTGTCAACAGTTTTTCTATGACTTTTAGTAGGATGAG 360
QY 994 AGCAGAAGTAGATCTCTTAAAGATGTTGAGAGGATCAAAATGTTTTTATATCAACATCC 1053
Db 361 AGCAGAAGTAGATCTCTTAAAGATGTTGAGAGGATCAAAATGTTTTTATATCAACATCC 420
QY 1054 TTTTATTATTGATTTCATTTGAGTTTAAACAGTGGTGTAGTGATAGATTTTCTATCTTTT 1113
Db 421 TTTTATTATTGATTTCATTTGAGTTTAAACAGTGGTGTAGTGATAGATTTTCTATCTTTT 480
QY 1114 CCCTTGACGTTTACTTTTCAAGTAACACAAACTCTCCATCAGGCCATGATCTATAGACC 1173
Db 481 CCCTTGACGTTTACTTTTCAAGTAACACAAACTCTCCATCAGGCCATGATCTATAGACC 540
QY 1174 TCCTAATAGAGATCTCTGGGTGATTTGTACCCCAACCATCTCTCCAAAGCATTAATATC 1233
Db 541 TCCTAATAGAGATCTCTGGGTGATTTGTACCCCAACCATCTCTCCAAAGCATTAATATC 600
QY 1234 CAATCATGCGCTGATGTTTAAATACGAGAAGCATGTTTATGTTGTACAAAAGAG 1293
Db 601 CAATCATGCGCTGATGTTTAAATACGAGAAGCATGTTTCCCATGTCCTGACAAAAGAG 660
QY 1294 ATTGTATTGGTGGGATGGAGGTAT---AGACCATGTCATGGTCACTTCAAG--CTACT 1348
Db 661 ATTGTATTGGTGGGATGGGCGCTCTCAGACCATGTCATGGTCACTTCAAGGCTACTT 720
QY 1349 TTAATAAAGGATCTTAAATGGCGAGGAGTGTGAACAAAGACACCCCTTAATATGGGTT 1408
Db 721 TTAATAAAGGATCTCAATGGCGAGGAGTGTGAACAAAGGAGTGTGAACAAAGGAGTGT 780
QY 1409 GATGTCCTGAAG-TAGCAAAATCTTCTGGAACGCAAACTCT 1447
Db 781 GATGTCCTGAAGTTAGCAAAATCTCTCTGGAACGCAAACTCT 820
```

## RESULT 8

BF975096

## LOCUS

BF975096 60245679F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:436722 5', DEFINITION mRNA sequence.



```

ACCESSION      BF975096
VERSION        1
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      1 (bases 1 to 693)
AUTHORS        NIH-MGC http://mgc.nci.nih.gov/
TITLE          National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL        Unpublished (1999)
COMMENT        Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-re@mail.nih.gov
                Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
                cDNA Library Preparation: Ling Hong/Rubin Laboratory
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLCMI209 row: n column: 19
                High quality sequence stop: 692.
                Location/Qualifiers
                1..693
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4336722"
                /clone_lib="NIH_MGC_48"
                /tissue_type="primary B-cells from tonsils (cell line)"
                /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                Site_2: EcoRI; cDNA made by oligo-dt priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Size-selected by Ling
                Hong in the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC Library."
                179 a 167 c 162 g 185 t
BASE COUNT
ORIGIN
Query Match      24.2%; Score 681; DB 12; Length 693;
Best Local Similarity 99.9%; Pred. No. 1.4e-70;
Matches 692; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 304 CTACCGCGTACCTGTTTCACTCTCTGGAGCCCTCTAGACTGTGCCGACATGTGCC 363
DB 1 CTACCGCGTACCTGTTTCACTCTCTGGAGCCCTCTAGACTGTGCCGACATGTGCC 60
QY 364 CGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTT 423
DB 61 CGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTT 120
QY 424 CTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGCGGCTGCACCGCGCGGGGTGCA 483
DB 121 CTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGCGGCTGCACCGCGCGGGGTGCA 180
QY 484 AATAGGCATCATGACCTTCAAGATTTATTTTACTGCTGGAATACTTTGTAGAAACCA 543
DB 181 AATAGGCATCATGACCTTCAAGATTTATTTTACTGCTGGAATACTTTGTAGAAACCA 240
QY 544 TGAAGAAGCTTTCAAAGCTGGGAAGGCTGCGATGAAATTCAGTTCGTCCTCCAGACA 603
DB 241 TGAAGAAGCTTTCAAAGCTGGGAAGGCTGCGATGAAATTCAGTTCGTCCTCCAGACA 300
QY 604 GCTTCGGCGCATCTTTTCCCTGTATGAGGTGATGACTTACGAGACGCAATTCGTGAC 663
DB 301 GCTTCGGCGCATCTTTTCCCTGTATGAGGTGATGACTTACGAGACGCAATTCGTGAC 360
QY 664 TTTGGGACTTTGATACCACTTCCAGGAATGTACACAGATGAATATCTCGTCGAAG 723
DB 361 TTTGGGACTTTGATACCACTTCCAGGAATGTACACAGATGAATATCTCGTCGAAG 420

```

---

```

QY 724 ACAGTGGATAAAAAACAGTCTCTCAAGTCTTCTCTGTTTATTTCTTCAACTCTCACTTT 783
DB 421 ACAGTGGATAAAAAACAGTCTCTCAAGTCTTCTCTGTTTATTTCTTCAACTCTCACTTT 480
QY 784 CTTAGAGTTTACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCATATGCTTTGAAA 843
DB 481 CTTAGAGTTTACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCATATGCTTTGAAA 540
QY 844 ATAGAGAAGAACACAGGCTGCGCCAGGAGGCTGCTGCAATTTGGTGAGTTTGGATGCA 903
DB 541 ATAGAGAAGAACACAGGCTGCGCCAGGAGGCTGCTGCAATTTGGTGAGTTTGGATGCA 599
QY 904 ACATTGTCCTCTACTGGGAATAACAGAACTCAGAGACTCGGAGCATCTTAAAGTGTCAA 963
DB 600 ACATTGTCCTCTACTGGGAATAACAGAACTCAGAGACTCGGAGCATCTTAAAGTGTCAA 659
QY 964 CGTTTTTCTATGACTTTTAGTAGGATGAGAGC 996
DB 660 CGTTTTTCTATGACTTTTAGTAGGATGAGAGC 692

RESULT 9
BF975166      942 bp      mRNA      linear      EST 22-JAN-2001
LOCUS        602244657F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4335639 5',
DEFINITION   mRNA sequence.
ACCESSION    BF975166
VERSION      BF975166.1 GI:12342381
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE    1 (bases 1 to 942)
AUTHORS      NIH-MGC http://mgc.nci.nih.gov/.
TITLE        National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL      Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-re@mail.nih.gov
                Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
                cDNA Library Preparation: Ling Hong/Rubin Laboratory
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                DNA Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLCMI207 row: a column: 16
                High quality sequence stop: 707.
                Location/Qualifiers
                1..942
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4335639"
                /clone_lib="NIH_MGC_48"
                /tissue_type="primary B-cells from tonsils (cell line)"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: B-cells; Vector: pOTB7; Site_1: XhoI;
                Site_2: EcoRI; cDNA made by oligo-dt priming.
                Directionally cloned into EcoRI/XhoI sites using the
                following 5' adaptor: GGCACGAG(G). Size-selected by Ling
                Hong in the laboratory of Gerald M. Rubin (University of
                California, Berkeley) using ZAP-cDNA synthesis kit
                (Stratagene) and Superscript II RT (Life Technologies).
                Note: this is a NIH_MGC Library."
                249 a 209 c 251 g 233 t
BASE COUNT
ORIGIN
Query Match      23.7%; Score 667.4; DB 12; Length 942;
Best Local Similarity 91.3%; Pred. No. 4.1e-69;
Matches 799; Conservative 0; Mismatches 61; Indels 15; Gaps 8;

```



Db 482 CCCTTGACGCTTACTTTCAAGTAACACAAACACTTCTCCATCAGGCCATGATCTATAGGACC 541

QY 1174 TCCTAATGAGAGTATCTGGGTGATTTGAGCCCCAAACCATCTCTCCAAAGCATTAATATC 1233

Db 542 TCCTAATGAGAGTATCTGGGTGATTTGAGCCCCAAACCATCTCTCCAAAGCATTAATATC 601

QY 1234 CAATCATGCGCTGATGTTTTAATCAGCAGAGACGATGTTTTTATGTTTGTACAAAAAGAAG 1293

Db 602 CAATCATGCGCTGATGTTTTAATCAGCAGAGACGATGTTTTTATGTTTGTACAAAAAGAAG 661

QY 1294 ATT 1296

Db 662 ATT 664

RESULT 11

BG757392

LOCUS 602711022F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4851580 5',

DEFINITION mRNA sequence.

ACCESSION BG757392

VERSION BG757392.1 GI:14068045

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 693)

TITLE NIH-MGC http://mgi.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLCMI694 row: k column: 05

High quality sequence stop: 693.

Location/Qualifiers

1. .693

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4851580"

/clone\_lib="NIH\_MGC\_48"

/tissue\_type="primary B-cells from tonsils (cell line)"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: B-cells; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH-MGC library."

BASE COUNT 157 a 188 c 178 g 170 t

ORIGIN

Query Match 23.3%; Score 655.4; DB 12; Length 693;

Best Local Similarity 99.7%; Pred. No. 1.3e-67;

Matches 667; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 AGAGAACCATCATTAATTGAAGTGAAGATTTTCTGGCCTGAGACTTTGCAGGGAGGCAAGA 60

Db 2 AGAGAACCATCATTAATTGAAGTGAAGATTTTCTGGCCTGAGACTTTGCAGGGAGGCAAGA 61

QY 61 AGACACTCTGGACACCACTATGACAGACCTCTTGATGACCGGAGGAAGTTTCTTTACCA 120

Db 62 AGACACTCTGGACACCACTATGACAGACCTCTTGATGACCGGAGGAAGTTTCTTTACCA 121

QY 121 ATTCAAAAATGTCGCTGGGCTAAGGCTCGCGCTGAGACCTACCTGTGCTACGTAGTGAA 180

Db 122 ATTCAAAAATGTCGCTGGGCTAAGGCTCGCGCTGAGACCTACCTGTGCTACGTAGTGAA 181

QY 181 -GAGGCGTGACAGTGTACATCCCTTTTCACTGGACTTTGGTTATCTTCGCAANTAGAAGC 239

Db 182 CGAGGGCTGACAGTGTACATCCCTTTTCACTGGACTTTGGTTATCTTCGCAANTAGAAGC 241

QY 240 GCTGCCACGCTGGGAATGCTCTTCCTCGCTACATCTCGGACTCGGACTCGGACCTAGACCTGGCC 299

Db 242 GCTGCCACGCTGGGAATGCTCTTCCTCGCTACATCTCGGACTCGGACTCGGACCTAGACCTGGCC 301

QY 300 GCTGCTACCGCGCTACCTGGTTTCACTCTCCCTGGAGCCCTGCTACGACTGTGCCGACATG 359

Db 302 GCTGCTACCGCGCTACCTGGTTTCACTCTCCCTGGAGCCCTGCTACGACTGTGCCGACATG 361

QY 360 TGGCCGACTTTTTCGAGAGGAAACCCCAACCTCAGCTCAGGATCTTCACCGCGGCTCT 419

Db 362 TGGCCGACTTTTTCGAGAGGAAACCCCAACCTCAGCTCAGGATCTTCACCGCGGCTCT 421

QY 420 ACTTCTGTGAGGACCGCAAGCTGAGCCCGAGGGGCTGCGGCGCTGCACCGCGCGGG 479

Db 422 ACTTCTGTGAGGACCGCAAGCTGAGCCCGAGGGGCTGCGGCGCTGCACCGCGCGGG 481

QY 480 TGCAATATAGCCATCATGACCTTCAAGATATATTTTACTGCTGGAATACTTTTGTAGAAA 539

Db 482 TGCAATATAGCCATCATGACCTTCAAGATATATTTTACTGCTGGAATACTTTTGTAGAAA 541

QY 540 ACCATGAAGAAGCTTTCAAGCCCTGGAAGGCTGCAATGAAAATTCAGTTCGTCTCTCA 599

Db 542 ACCATGAAGAAGCTTTCAAGCCCTGGAAGGCTGCAATGAAAATTCAGTTCGTCTCTCA 601

QY 600 GACAGCTTCGCGCATCCTTTTGCCTGTATGAGGTTGATGACTTAGGACGACCATTC 659

Db 602 GACAGCTTCGCGCATCCTTTTGCCTGTATGAGGTTGATGACTTAGGACGACCATTC 661

QY 660 GTACTTTGG 668

Db 662 GTACTTTGG 670

RESULT 12

BG341546

LOCUS 602463652F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4576548 5',

DEFINITION mRNA sequence.

ACCESSION BG341546

VERSION BG341546.1 GI:13147984

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS 1 (bases 1 to 670)

TITLE NIH-MGC http://mgi.nci.nih.gov/.

JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)

COMMENT Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: Ling Hong/Rubin Laboratory

CDNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLCMI289 row: g column: 13

High quality sequence stop: 650.

Location/Qualifiers

1. .670

/organism="Homo sapiens"

/db\_xref="taxon:9606"

FEATURES

source

/clone="IMAGE:4576548"  
/clone\_lib="NIH\_MGC\_48"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."  
200 a 122 c 134 g 214 t

Query Match 22.3%; Score 627.6; DB 12; Length 670;  
Best Local Similarity 98.8%; Pred. No. 2.3e-64;  
Matches 664; Conservative 0; Mismatches 4; Indels 4; Gaps 3;  
QY 634 GGTGTGACTTACGAGACGCAATTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAAT 693  
DB 2 GGTGTGACTTACGAGACGCAATTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAAT 60  
QY 694 GTCACACGATGAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCCTTCAAGTCT 753  
DB 61 GTCACACGATGAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCCTTCAAGTCT 120  
QY 754 TCTCTGTTTTTATCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAAAATATTTATATA 813  
DB 121 TCTCTGTTTTTATCTTCAACTCTCACTTTCTTAGAGTTTACAGAAAAAATATTTATATA 180  
QY 814 CGACTCTTTAAAAAGATCTATGCTTGAATAAGAGAACACACAGGTCGCGCCAGGGA 873  
DB 181 CGACTCTTTAAAAAGATCTATGCTTGAATAAGAGAACACACAGGTCGCGCCAGGGA 240  
QY 874 CGTCTGCAATTCGGTCAGTTTGAATGCAACATTTGCCCTACTGGGAAATACAGAACT 933  
DB 241 CGTCTGCAATTCGGTCAGTTTGAATGCAACATTTGCCCTACTGGGAAATACAGAACT 300  
QY 934 GCAGGACCTGGGAGCATCCTTAAAGTGTCAACGTTTTTCTATGACTTTTAGGTAGGATGAG 993  
DB 301 GCAGGACCTGGGAGCATCCTTAAAGTGTCAACGTTTTTCTATGACTTTTAGGTAGGATGAG 360  
QY 994 AGCAGAAGTGTAGTCTTAAAGATGTTGAGAGGATCAAAATGTTTTTATATCAACATCC 1053  
DB 361 AGCAGAAGTGTAGTCTTAAAGATGTTGAGAGGATCAAAATGTTTTTATATCAACATCC 420  
QY 1054 TTTATTTATTTGATTCATTTGAGTTAACAGTGGTGTAGTATAGATTTTCTTATCTTTT 1113  
DB 421 TTTATTTATTTGATTCATTTGAGTTAACAGTGGTGTAGTATAGATTTTCTTATCTTTT 480  
QY 1114 CCCTTGACGTTTACTTTCAAGTAACACAACTCTCCATCAGGCCATGATCTATAGACC 1173  
DB 481 CCCTTGACGTTTACTTTCAAGTAACACAACTCTCCATCAGGCCATGATCTATAGACC 540  
QY 1174 TCCTAATGAGATATCGGTGATTTGACCCCAACCACTCTCTCC- AAAGCATTAATAT 1232  
DB 541 TCCTAATGAGATATCGGTGATTTGACCCCAACCACTCTCTCCAAAGCATTAATAT 600  
QY 1233 CCAATCATGGCTGTATGTTTTTAATCAGCAGAAGCATGTTTTTATGTTGTACAAAGAA 1292  
DB 601 CCAATCATGGCTGTATGTTTTTAATCAGCAGAAGCATGTTTTTATGTTGTACAAAGAA 658  
QY 1293 GATTGTTATGGG 1304  
DB 659 GATTGTTATGGG 670

RESULT 13  
BF664355 956 bp mRNA linear EST 21-DEC-2000  
LOCUS 602146013F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4309707 5',  
DEFINITION

mRNA sequence.  
BF664355  
BF664355.1 GI:11938351  
EST.  
Homo sapiens  
human.

ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 956)  
NIH-MGC http://mgs.nci.nih.gov/.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-r@mail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1183 row: i column: 04  
High quality sequence stop: 725.

FEATURES  
source

1..956  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4309707"  
/clone\_lib="NIH\_MGC\_48"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."  
277 a 183 c 219 g 277 t

BASE COUNT  
ORIGIN

Query Match 22.2%; Score 626; DB 12; Length 956;  
Best Local Similarity 97.2%; Pred. No. 2.6e-64;  
Matches 690; Conservative 0; Mismatches 15; Indels 5; Gaps 5;  
QY 634 GGTGTGACTTACGAGACGCAATTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAAT 693  
DB 2 GGTGTGACTTACGAGACGCAATTCGTACTTTGGGACTTTGATAGCAACTTCCAGGAAT 60  
QY 694 GTCACACGATGAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCCTTCAA-GTC 752  
DB 61 GTCACACGATGAATATCTCTGCTGAAGACAGTGGATAAAAAACAGTCCTTCAAAGTC 120  
QY 753 TTTCTGTTTTTATTTTCAACTCTCACATTTCTTAGAGTTTACAGAAAAAATATTTATAT 812  
DB 121 TTTCTGTTTTTATTTTCAACTCTCACATTTCTTAGAGTTTACAGAAAAAATATTTATAT 180  
QY 813 AGCAGTCTTTAAAGATCTATGCTTTGAAAATAGAGGAACACAGGTCGCGCCAGG 872  
DB 181 AGCAGTCTTTAAAGATCTATGCTTTGAAAATAGAGGAACACAGGTCGCGCCAGG 240  
QY 873 AGTGTGCTCAATTTGGTGCAGTTTGAATGCAACATTTGCCCTACTGGGAATAACAGAAC 932  
DB 241 AGTGTGCTCAATTTGGTGCAGTTTGAATGCAACATTTGCCCTACTGGGAATAACAGAAC 300  
QY 933 TGCAGGACCTGGGAGCATCTTAAAGTGTCAACGTTTTTCTATGACTTTTAGGTAGGATGA 992  
DB 301 TGCAGGACCTGGGAGCATCTTAAAGTGTCAACGTTTTTCTATGACTTTTAGGTAGGATGA 360  
QY 993 GAGCAGAAGTGTAGTCTTAAAGATGTTGAGAGGATCAAAATGTTTTTATATCAACATC 1052  
|||||

```
Db 361 GAGCAGAGGTAGATCTCTAAAGCATGGTGAGAGATCAAAATGTTTTATATCAACATC 420
QY 1053 CTTTATTATTGATTCATT -TGAGTTAACAGTGGTGTAGTGATAGATTTTCTATTCTT 1111
Db 421 CTTTATTATTGATTCATTGTGAGTTAACAGTGGTGTAGTGATAGATTTTCTATTCTT 480
QY 1112 TTCCCTTGAGCTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCATGATCTATAGGA 1171
Db 481 TTCCCTTGAGCTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCATGATCTATAGGA 540
QY 1172 CCTCCTAATGAGAGTATCTGGGTGATGTGACCCCAACCATCTCTCCAAAGCATTAATA 1231
Db 541 CCTCCTAATGAGAGTATCTGGGTGATGTGACCCCAACCATCTCTCCAAAGCATTAATA 599
QY 1232 TCCAATCATCGCGCTGATGTTTTAATCAGCAGAGCATGTTTTATGTTTGTACAAAAGA 1291
Db 600 TCCATCATCGCGCTGATGTTTTAATCAGCAGAGCATGTTTTATGTTTGTACAAAAGA 658
QY 1292 AGATTGTTTATGGGTGGGATGGAGGTATAGACCATGATGATGATGATGATGATGATGAT 1341
Db 659 AGATTGTTTATGGGTGGGATGGAGGTATAGACCATGATGATGATGATGATGATGATGAT 708

RESULT 14
AW978582/c
LOCUS 604 bp mRNA linear EST 02-JUN-2000
DEFINITION EST390691 MAGE resequences, MAGP Homo sapiens cDNA, mRNA sequence.
ACCESSION AW978582
VERSION AW978582.1 GI:8169850
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 604)
AUTHORS Hegde,P., Qi,R., Abernathy,K., Dharap,S., Gaspard,R., Gay,C., Hoit
,I.E., Saeed,A.I., Sharov,V., Lee,N.H., Yeatman,T.J. and
Quackenbush,J.
TITLE Assessment of gene expression patterns in a model of colon tumor
metastasis using a 19,200 element cDNA microarray
JOURNAL Unpublished (2000)
COMMENT Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 3528
Fax: 301 838 0208
Email: johng@tigr.org
Plate: 394
Seq primer: Forward.
Location/Qualifiers
1..604
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="MAGE resequences, MAGP"
/note="Vector: pBluescriptSkm"
BASE COUNT 190 a 121 c 109 g 184 t
ORIGIN

Query Match 20.3%; Score 572.6; DB 10; Length 604;
Best Local Similarity 97.6%; Pred. No. 6e-58;
Matches 581; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 1061 TTTGATTCATTTGAGTTAAACAGTGGTGTAGTGATAGATTTTCTATTCTTTTCCCTTGA 1120
Db 595 TTGAATCCATTTGAGTTAAACAGTGGTGTAGTGATAGATTTTCTATTCTTTTCCCTTGA 536
QY 1121 CTTTACTTCAAGTAACACAAACTCTCCATCAGGCCATGATCTATAGGACCTCCTAAT 1180
Db 535 CTTTATTATTCAAGTAACACAAACTCTCCATCAGGCCATGTTTATAGACCTCCTAAT 476
QY 1181 GAGAGTATCTGGGTGATGTGACCCCAACCATCTCTCCAAAGCATTAATATCAATCAT 1240
Db 475 GAGAGTATCTGGGTGATGTGACCCCAACCATCTCTCCAAAGCATTAATATCAATCAT 416
```

```
QY 1241 GCCTGTATGTTTTAATCAGCAGAGCATGTTTTATGTTTGTACAAAAGAGATTGTTA 1300
Db 415 GCCTGTATGTTTTAATCAGCAGAGCATGTTTTATGTTTGTACAAAAGAGATTGTTA 356
QY 1301 TGGGTGGGATGGAGGTATAGACCATGATGGTCCACCTTCAAGCTACTTTTAATAAGGAT 1360
Db 355 TGGGTGGGATGGAGGTATAGACCATGATGGTCCACCTTCAAGCTACTTTTAATAAGGAT 296
QY 1361 CTTAAAATGGCAGGAGACTGTGAACAAGACACCCCTAATAATGGTGTGATCTCTGAAGT 1420
Db 295 CTTAAAATGGCAGGAGACTGTGAACAAGACACCCCTAATAATGGTGTGATCTCTGAAGT 236
QY 1421 AGCAAAATCTTCTGAAAACGCAAACTCTTTTAAAGGAAGTCCCTAATTTAGAACAACCCACA 1480
Db 235 AGCAAAATCTTCTGAAAACGCAAACTCTTTAAAGGAAGTCCCTAATTTAGAACAACCCACA 176
QY 1481 AACTTCACATATCATATAATTAGCAAAACAATTGGAAGGAAGTTCCTTGAATTTGGGGAGAG 1540
Db 175 AACTTCACATATCATATAATTAGCAAAACAATTGGAAGGAAGTTCCTTGAATTTGGGGAGAG 116
QY 1541 GAAATATCTATTGGCTCTCGTGGGTCTCTTCACTCAGAAATGCCAATCAGGTCAAGGTTT 1600
Db 115 GAAATATCTATTGGCTCTCGTGGGTCTCTTCACTCAGAAATGCCAATCAGGTCAAGGTTT 56
QY 1601 GCTACATTTTGTATGTGTGATGCTTCTCCCAAGGTATATTAACCTATATAAGA 1655
Db 55 GCTACATTTTGTATGTGTGATGCTTCTCCCAAGGTATATTAACCTATATAANTA 1

RESULT 15
BG398364
LOCUS 928 bp mRNA linear EST 12-MAR-2001
DEFINITION 60243983ZF1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4566366 5',
mRNA sequence.
ACCESSION BG398364
VERSION BG398364.1 GI:13291812
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 928)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1284 row: o column: 07
High quality sequence stop: 650.
Location/Qualifiers
1..928
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4566366"
/tissue_type="primary B-cells from tonsils (cell line)"
/note="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
FEATURES
source
```

Note: this is a NIH\_MGC Library."

BASE COUNT	280 a	163 c	222 g	263 t	
ORIGIN					
Query Match	20.2%	Score 568;	DB 12;	Length 928;	
Best Local Similarity	98.6%	Pred. No. 1.5e-57;			
Matches 615; Conservative	0;	Mismatches 5;	Indels 4;	Gaps 4;	
Qy	622	GGCCCTGTATGAGGTTGATGACATTACGAGACGCTTTCGTACTTTGGGACTTTGATAGCA	681		
Db	55	GGCCCTGTATGAGGTTGATGACATTACGAGACGCTTTCGTACTTTGGGACTTTGATAGCA	113		
Qy	682	ACTTCCAGGAATGCACACGATGAAATATCTCTGCTGAAGACAGTGGATAAAAAACAG	741		
Db	114	ACTTCCAGGAATGCACACGATGAAATATCTCTGCTGAAGACAGTGGATAAAAAACAG	173		
Qy	742	TCCTTCAAGTCCTCTCTGTTTTTATCTTCAACTCTCAGTTTCTTAGAGTTTACAGAAA	801		
Db	174	TCCTTCAAGTCCTCTCTGTTTTTATCTTCAACTCTCAGTTTCTTAGAGTTTACAGAAA	233		
Qy	802	AATATTATATAGGACTCTTTAAAAAGATCTATGCTTTGAAAAATAGAGAGGAACACAGG	861		
Db	234	AATATTATATAGGACTCTTTAAAAAGATCTATGCTTTGAAAAATAGAGAGGAACACAGG	293		
Qy	862	TCGGCCAGGAGCTGCTGCAATTGGTGCAGTTTTTGAATGCAACATTGCCCCCTACTGGG	921		
Db	294	TCGGCCCA-GGACGTGCTGCAATTGGTGCAGTTTTTGAATGCAACATTGCCCCCTACTGGG	352		
Qy	922	AATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGTCACGTTTTTCTATGACTTTT	981		
Db	353	AATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGTCACGTTTTTCTATGACTTTT	412		
Qy	982	AGGTAGGATGAGAGCAGAGGTAGATCCTTAAAAAGCATGCTGAGAGGATCAAATG-TTTT	1040		
Db	413	AGGTAGGATGAGAGCAGAGGTAGATCCTTAAAAAGCATGCTGAGAGGATCAAATGTTT	472		
Qy	1041	TATATCAACATCCTTTTATTATTGATTCATTTGAGTTAACAGTGGTGTAGTAGATAGATT	1100		
Db	473	TATATCAACATCCTTTTATTATTGATTCATTTGAGTTAACAGTGGTGTAGTAGATAGATT	532		
Qy	1101	TTTCTATTCTTTCCCTTGACGTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCAT	1160		
Db	533	TTTCTATTCTTTCCCTTGACGTTTACTTTCAAGTAACACAAACTCTTCCATCAGGCCAT	592		
Qy	1161	GATCTATAGGACCTCCTTAATGAGAGTATCTGGGTGATTGTGACCCCAAAACATCTCTCCA	1220		
Db	593	GATCTATAGGACCTCCTTAATGAGAGTATCTGGGTGATTGTGACCCCAAA-CATCTCTCCA	651		
Qy	1221	AAGCATTAATATCAATCATGCGC	1244		
Db	652	AAGCATTAATATCAATCATGCGC	675		

Search completed: June 19, 2003, 11:07:25  
Job time : 3805.83 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:01:09 ; Search time 118.35 Seconds

(without alignments)

7302.178 Million cell updates/sec

Title: US-09-966-880A-7

Perfect score: 2818

Sequence: 1 agagaccattcattatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*

1: /cgn2.6/ptodata/1/ina/5A-COMB.seq:\*

2: /cgn2.6/ptodata/1/ina/5B-COMB.seq:\*

3: /cgn2.6/ptodata/1/ina/6A-COMB.seq:\*

4: /cgn2.6/ptodata/1/ina/6B-COMB.seq:\*

5: /cgn2.6/ptodata/1/ina/PTUS-COMB.seq:\*

6: /cgn2.6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	280	9.9	15297	4	US-09-817-180-3
2	274.2	9.7	11811	4	US-09-078-294-7
3	269.8	9.6	43950	4	US-09-735-934A-3
4	269.2	9.6	99500	4	US-09-798-096-10
5	267.4	9.5	59065	4	US-09-813-817-3
6	267.4	9.5	59065	4	US-09-978-197-3
7	265.6	9.4	9365	4	US-09-608-285A-8
8	265.6	9.4	9365	4	US-09-350-836B-8
9	265.6	9.4	9365	4	US-09-370-265-8
10	265.6	9.4	14747	4	US-09-608-285A-42
11	265.6	9.4	15977	4	US-09-608-285A-59
12	262.2	9.3	21234	4	US-09-810-671-3
13	259.2	9.2	3885	1	US-08-688-145-1
14	258.6	9.2	43950	4	US-09-735-934A-3
15	257.6	9.1	99500	4	US-09-798-096-10
16	257.2	9.1	10684	3	US-08-618-100B-3
17	256.8	9.1	8174	1	US-07-914-281-5
18	256.8	9.1	8174	1	US-08-393-246-5
19	256.8	9.1	8174	1	US-08-525-058A-5
20	256.8	9.1	8174	2	US-08-696-731-5
21	256.8	9.1	8174	4	US-09-042-531-5
22	256.8	9.1	8174	5	PCT-US91-00899-3
23	256.8	9.1	29629	4	US-09-729-995-3
24	254.4	9.0	98844	4	US-09-791-211-10
25	253.8	9.0	16063	4	US-09-801-052-3
26	253.2	9.0	14581	4	US-08-520-373D-4
27	253.2	9.0	22481	4	US-08-367-841A-43

ALIGNMENTS

RESULT 1

US-09-817-180-3

; Sequence 3, Application US/09817180

; Patent No. 6340584

; GENERAL INFORMATION:

; APPLICANT: GAN, Weinlu et al.

; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC

; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES

; FILE REFERENCE: CLO01183

; CURRENT APPLICATION NUMBER: US/09/817,180

; CURRENT FILING DATE: 2001-03-27

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 3

; LENGTH: 15297

; TYPE: DNA

; ORGANISM: Human

US-09-817-180-3

Query Match 9.9%; Score 280; DB 4; Length 15297;

Best Local Similarity 74.9%; Pred. No. 3.9e-48;

Matches 370; Conservative 0; Mismatches 110; Indels 14; Gaps 1;

QY	1641	ATTAACATATATAGAGAGTTGTGACAAACAGAAATGATAAGCTGGCAACCGTGGCACAC	1700
Db	11651	ACTCCCTTAATGCCAACCTTCCACCAGGAGAGTAAGAAATACCTGGCCAGTTGCTCAC	11710
QY	1701	GCTCATAGTCTTACGTCTTGGAGGTTGAGGAGGAGGATGCTTGAACACAGGTGTTTC	1760
Db	11711	GCCTGTTCATCCAGCACTTTGGGAGGCTGAGCTGGGTGATCATCTTGAGCCAGGAGTTTC	11770
QY	1761	AAGCCAGCGCTGGGCAACATACAAAGATCCTGCTCAAAAAAATAAATAAATAAATAA	1820
Db	11771	AGATCAGCTTGGCAACACAGTGAACCTCATCTGTACAAAAATACAAAAATA	11825
QY	1821	AAGAGAGAGCGCGCGGTGGTGCCTCACCGCTGTATCCAGCACTTTTGGGAGCGCGAG	1880
Db	11826	-----GACTGGGCACGGTGGCTCACACCTGTATCCAGCACTTTTGGGAGCGCGAG	11876
QY	1881	CGGGGCGGATCAGCTGTGGTCAGGAGTTTGAGACCAAGCTGGCCACATGGGCAAAACCCC	1940
Db	11877	GCAGTGGATCAGCTGTGGTCAGGAGTTTGAGACCAAGCTGGCCACATGGTGAACACCCC	11936
QY	1941	GTCGTACTCAAAATGCAAAATAGCCAGCGGTGTAGCAGGCACCTGTAATCCCAAGCT	2000
Db	11937	ATCTTACATAAAATACAAAAATAGCCAGGATGGTGGCACGCTGCTGTAATCCCAAGCT	11996
QY	2001	ACTTGGGAGCTGAGGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGTTGCAGTAGCT	2060

```

Db      11997 ACTTGGAGGCTGAGGTGGGAGAAATTCCTTGAACCCAGGAGCGGAGGCTGCAGTGAGCC 12056
Qy      2061 GAGATCTGCTGGTTCGACTCCAGCTGGGCGGACAGCAAGACTCTGTCTCAGAAAAA 2120
Db      12057 GAGATTGTGCGCACTGCACTCCAGCTCGGCGGACAGAGTGAACCTCATCTCAAAAAA 12116
Qy      2121 AAAAAAAGAGAGA 2134
Db      12117 CCAAAAAACAAAA 12130

RESULT 2
US-09-078-294-7/c
; Sequence 7, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 11811
; TYPE: DNA
; ORGANISM: BAC-F2 contig 3
US-09-078-294-7

Query Match          9.7%; Score 274.2; DB 4; Length 11811;
Best Local Similarity 79.7%; Pred. No. 5.5e-47;
Matches 350; Conservative 0; Mismatches 83; Indels 6; Gaps 2;

Qy      1692 GTGGCACAGCTCATAGTTCTAGCTGTGGAGTTTGAGGAGGAGGAGTGGCTTGAACA 1751
Db      6663 GTGGCTCATCTGTATATCCAGCACTTTGGAGCGGAGGTGGAGAGTGAAGTGGCT 6604
Qy      1752 CAGGTGTTCAAGGCGAGCTGGGCAACATACACAGATCTCTCTCAAAAAA 1811
Db      6603 CAGAAGTTCAGAGCAGCTGGCCACACAGTGAAGCTCATCTCTACAAAATACATTTA 6544
Qy      1812 AAAAAAGAAAGAGAGAGGCGGCGTGTGGTGTACGCTGTATATCCAGCACTTTGG 1871
Db      6543 AAGCTTAG- ----CTGGGGCCAGGTGTGTGGCGCAGCTGTATATCCAGCCCTTTGG 6489
Qy      1872 GAGCGGAGCGGCGGATCACCTGTGTCAGGAGTTTGAGACAGCCTGGCCCAACATGG 1931
Db      6488 GAGTCAAGTGTGGTGGATCACCTGAGTTCAGGAGTTTCGAGCCAGCCTGGCCCAACATGG 6429
Qy      1932 CAAAAACCGTCTGTACTCAAAATGC- AAAAAATTAGCCAGGCGTGTAGCAGGCACCTGT 1990
Db      6428 TGAACCCCATCTCTACTAATAATACAAAAATTAGCCAGGCGTGTGGCGGCGCTGT 6369
Qy      1991 AATCCAGCTACTTGGAGGCTGAGGAGGAGTTCCTTGAACCCAGGAGTGGAGGTT 2050
Db      6368 AATCCAGCTACTTGGAGGCTGAGGAGGAGTTCCTTGAACCCAGGAGGAGT 6309
Qy      2051 GCAGTAACTGAGATGTCGCTGTGACCTCCAGCTGGCGGACAGCAAGCACTGTTC 2110
Db      6308 GCAGTAACTGAGATGTCGCTGTGACCTCCAGCTGGCGGACAGCAAGCACTGTTC 6249
Qy      2111 TCAGAAAAA 2129
Db      6248 AAAAAA 6230

RESULT 3
US-09-735-934A-3/c
; Sequence 3, Application US/09735934A
; Patent No. 6372468
; GENERAL INFORMATION:

```

```

; APPLICANT: Li, Jiayin et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CLO00851
; CURRENT APPLICATION NUMBER: US/09/735,934A
; CURRENT FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 43950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-735-934A-3

Query Match          9.6%; Score 269.8; DB 4; Length 43950;
Best Local Similarity 77.9%; Pred. No. 6.1e-46;
Matches 339; Conservative 0; Mismatches 92; Indels 4; Gaps 1;

Qy      1685 GCGAACCGTGGCACACGCTCATAGTTCTTCTAGCTGCTGGGAGGTTGAGGAGGAGTGGC 1744
Db      17761 GGGCACAGTGGCACACACCTGTATCCAGAACTTTGGGAGGCTGAGGTGGAGGATTC 17702
Qy      1745 TTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATAACAAGATCTCTCTCAAAAAA 1804
Db      17701 TTGAGCGCAGAAGTTCAAGACCACCTGGGCAACATAGCAAGACCCCTTATCTTAAAAA 17642
Qy      1805 AAAAAA 1860
Db      17641 AAAAAA 17582
Qy      1861 CAGCACTTTGGAGGCGGAGCGGCGGATACCTGTGTGTCAGGAGTTTGAGACCACT 1920
Db      17581 CAGCACTTTGGAGGCGGAGCGGAGTGTACCTGAGGTGGAGTTTCAAGACCACT 17522
Qy      1921 GGCACATGCGCAAAACCCCGTGTCTACTCAAAATGCAAAATTTAGCCAGGCGTGTAGC 1980
Db      17521 GACCAACATGGCAAAACCCCTGTCTCTACTAAAAATACAAAAAGTAGCTGGCGCTGGGC 17462
Qy      1981 AGCACTCTGTAAATCCAGCTACTTTGGAGGCTGAGCGAGAGAGTGGCTTGAACCCAGGA 2040
Db      17461 GGTGCGCTGTAGTCCAGCTACTTTGGAGGCTGAGCGAGAGAGTGGCTTGAACCCAGGA 17402
Qy      2041 GGTGAGGTTGAGTAAAGTGAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
Db      17401 GACAGAGTTGCAATGAACTGAAATGTCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 17342
Qy      2101 AGACTCTGCTCAGA 2115
Db      17341 AAACGCGCTCAAAA 17327

RESULT 4
US-09-798-096-10/c
; Sequence 10, Application US/09798096
; Patent No. 6399378
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF RECQL2 EXPRESSION
; FILE REFERENCE: RTS-0207
; CURRENT APPLICATION NUMBER: US/09/798,096
; CURRENT FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 10
; LENGTH: 99500
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-798-096-10

Query Match          9.6%; Score 269.2; DB 4; Length 99500;
Best Local Similarity 73.7%; Pred. No. 1e-45;

```



[illegible]



```

; NAME/KEY: misc_feature
; LOCATION: (9214)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymine
; NAME/KEY: misc_feature
; LOCATION: (9303)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymine
; NAME/KEY: misc_feature
; LOCATION: (9311)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymine
US-09-350-836B-8

Query Match          9.4%; Score 265.6; DB 4; Length 9365;
Best Local Similarity 79.0%; Pred. No. 2.9e-45;
Matches 297; Conservative 34; Mismatches 34; Indels 11; Gaps 2;

QY 1765 CCAGCTGGGCAACATAACAAGATCCTGCTCTCAAAAAAAAAAAAAAAAAAG- 1823
Db 7802 CCAGCATGGGTGACAGCAAGACCCCTGCTCAAAAAAAAAAAAAAAAAAG 7743

QY 1824 -----AGAGAGGCCGGCGTGGTGTGCTCAGCGCTGTAATCCAGCACCTTTGGGAGG 1875
Db 7742 AAMCGAMCAAAAAAGGCGTGGTGTGCTCAYRCCTRTAATCYAGCACCTTTGGGAGG 7683

QY 1876 CCGAGCGGGCGGATCACCTGTGGTCAGGAGTTTGAGACAGCCTGGCCCAACATGCAAA 1935
Db 7682 CYGAGGYGGGTGGATCACCTGAGRTCAGRAGTTTTRAGAMYAGCCTGCCCAACATGCTGAA 7623

QY 1936 ACCCGCTCTGTACTCAAAATGC--AAAAATTAGCCAGGCGTGTAGCAGCACCTGTAAT 1993
Db 7622 ACCCGCTCTGTACTCAAAATGC--AAAAATTAGCCAGGCGTGTGGCGCGTGTGCTGTAAT 7563

QY 1994 CCCAGCTTGGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2053
Db 7562 CCCAGCTTGGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7503

QY 2054 GTAAGCTGAGTGTGCGGCTGTGCTGCTCCAGCTGGGCGGACAGCAAGACTGTGCTCA 2113
Db 7502 GTGAGCYGNGATCRYCAYTGTGCTCCAGCTCCAGCCYRGMAACAGAGCAGACTCCGCTCA 7443

QY 2114 GAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 2129
Db 7442 AAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7427
```

```

RESULT 9
US-09-370-265-8/c
; Sequence 8, Application US/09370265
; Patent No. 644771
; GENERAL INFORMATION:
; APPLICANT: Mulero, Julio
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL CD39-LIKE
; TITLE OF INVENTION: POLYPEPTIDES
; FILE REFERENCE: 2811/35908
; CURRENT APPLICATION NUMBER: US/09/370,265
; EARLIER FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: PCT/US99/16180
; EARLIER FILING DATE: 1999-07-16
; EARLIER APPLICATION NUMBER: 09/350,836
; EARLIER FILING DATE: 1999-07-09
; EARLIER APPLICATION NUMBER: 09/273,447
; EARLIER FILING DATE: 1999-03-19
; EARLIER APPLICATION NUMBER: 09/244,444
; EARLIER FILING DATE: 1999-02-04
; EARLIER APPLICATION NUMBER: 09/122,449
; EARLIER FILING DATE: 1998-07-24
; EARLIER APPLICATION NUMBER: 09/118,205
; EARLIER FILING DATE: 1998-07-16
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 9365
; TYPE: DNA
```

```

; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (72)..(287)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1280)..(1579)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1819)..(1854)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2466)..(2555)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2863)..(2940)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3887)..(3952)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (4896)..(4994)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5846)..(5986)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (6965)..(7138)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (8556)..(8639)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3409)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymine
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9214)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymine
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9303)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymine
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9311)
; OTHER INFORMATION: n = adenine or guanine or cytosine or thymine
US-09-370-265-8

Query Match          9.4%; Score 265.6; DB 4; Length 9365;
Best Local Similarity 79.0%; Pred. No. 2.9e-45;
Matches 297; Conservative 34; Mismatches 34; Indels 11; Gaps 2;

QY 1765 CCAGCTGGGCAACATAACAAGATCCTGCTCTCAAAAAAAAAAAAAAAAAAG- 1823
Db 7802 CCAGCATGGGTGACAGCAAGACCCCTGCTCAAAAAAAAAAAAAAAAAAG 7743

QY 1824 -----AGAGAGGCCGGCGTGGTGTGCTCAGCGCTGTAATCCAGCACCTTTGGGAGG 1875
Db 7742 AAMCGAMCAAAAAAGGCGTGGTGTGCTCAYRCCTRTAATCYAGCACCTTTGGGAGG 7683

QY 1876 CCGAGCGGGCGGATCACCTGTGGTCAGGAGTTTGAGACAGCCTGGCCCAACATGCAAA 1935
Db 7682 CYGAGGYGGGTGGATCACCTGAGRTCAGRAGTTTTRAGAMYAGCCTGCCCAACATGCTGAA 7623

QY 1936 ACCCGCTCTGTACTCAAAATGC--AAAAATTAGCCAGGCGTGTAGCAGCACCTGTAAT 1993
Db 7622 ACCCGCTCTGTACTCAAAATGC--AAAAATTAGCCAGGCGTGTGGCGCGTGTGCTGTAAT 7563

QY 1994 CCCAGCTTGGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2053
Db 7562 CCCAGCTTGGGAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 7503
```

QY 2054 GTRAGCTGAGATCGTCCGCTTGCACCTCCAGCCTGGCGACAGAGCAAGAACTCTGTCTCA 2113  
 Db 7502 GTGAGCYGAGATCGTCCGCTTGCACCTCCAGCCTGGCGACAGAGCAAGAACTCCGTCTCA 7443  
 QY 2114 GAAAAAAGAAAAA 2129  
 Db 7442 AAAAAATAATAATA 7427

RESULT 10  
 US-09-608-285A-42/c  
 ; Sequence 42, Application US/09608285A  
 ; Patent No. 6335013  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ford, John  
 ; APPLICANT: Mulero, Julio  
 ; APPLICANT: Yeung, George  
 ; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE  
 ; FILE REFERENCE: 28110/36570  
 ; CURRENT APPLICATION NUMBER: US/09/608,285A  
 ; CURRENT FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: 09/583,231  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 09/557,800  
 ; PRIOR FILING DATE: 2000-04-25  
 ; PRIOR APPLICATION NUMBER: 09/481,238  
 ; PRIOR FILING DATE: 2000-01-11  
 ; PRIOR APPLICATION NUMBER: 09/370,265  
 ; PRIOR FILING DATE: 1999-08-09  
 ; PRIOR APPLICATION NUMBER: PCT/US99/16180  
 ; PRIOR FILING DATE: 1999-07-16  
 ; PRIOR APPLICATION NUMBER: 09/350,836  
 ; PRIOR FILING DATE: 1999-07-09  
 ; PRIOR APPLICATION NUMBER: 09/273,447  
 ; PRIOR FILING DATE: 1999-03-19  
 ; PRIOR APPLICATION NUMBER: 09/244,444  
 ; PRIOR FILING DATE: 1999-02-04  
 ; PRIOR APPLICATION NUMBER: 09/122,449  
 ; PRIOR FILING DATE: 1998-07-24  
 ; PRIOR APPLICATION NUMBER: 09/118,205  
 ; PRIOR FILING DATE: 1998-07-16  
 ; NUMBER OF SEQ ID NOS: 60  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 42  
 ; LENGTH: 14747  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc\_feature  
 ; LOCATION: (13641)  
 ; OTHER INFORMATION: n = adenosine or guanine or cytosine or thymidine  
 US-09-608-285A-42

Query Match 9.4%; Score 265.6; DB 4; Length 14747;  
 Best Local Similarity 79.08; Pred. No. 3.3e-45;  
 Matches 297; Conservative 34; Mismatches 34; Indels 11; Gaps 2;  
 QY 1765 CCAGCCTGGCAACATACAGATCTCTCTCAAAAAAAAAAAAAAAAAAGAG- 1823  
 Db 11162 CCAGCATGGGTGACAGAGCAAGCCTCTCAAAAAAAAAAAAAAAAAAG 1103  
 QY 1824 -----AGAGAGGCGCGCTGGTGGCTACGCTGTATCCAGCACTTTGGAGG 1875  
 Db 11102 AAMCGAMCAAAAAAGCYRGGMGTGGTCTCAIRCTTATCTTATCTTGGAGG 11043  
 QY 1876 CCAGCGGCGGATACCTGTGGTCAGGATTTGACAGCCTGGCCAAACATGCAAA 1935  
 Db 11042 CYGAGGCGGCGGATACCTGTGGTCAGGATTTGACAGCCTGGCCAAACATGTA 10983  
 QY 1936 ACCCGCTCTGTACTCAAAATGC-AAAAATAGCCAGGCGGTAGCAGGCGACCTGTAA 1993  
 Db 10982 ACCCGCTCTGTACTCAAAATGC-AAAAATAGCCAGGCGGTAGCAGGCGACCTGTAA 10923

QY 1994 CCCAGCTACTTGGGAGGCTGAGGAGAGAAATCGTTGAACCCAGGAGGTGAGGTTGCA 2053  
 Db 10922 CCCAGCTACTTGGGAGGCTGAGGAGAGAAATCGTTGAACCCAGGAGGTGAGGTTGCA 10863  
 QY 2054 CTAAGCTGAGATCGTCCGCTTGCACCTCCAGCCTGGCGACAGAGCAAGAACTCTGTCTCA 2113  
 Db 10862 GTGAGCYGAGATCGTCCGCTTGCACCTCCAGCCTGGCGACAGAGCAAGAACTCCGTCTCA 10803  
 QY 2114 GAAAAAAGAAAAA 2129  
 Db 10802 AAAAAATAATAATA 10787

RESULT 11  
 US-09-608-285A-59/c  
 ; Sequence 59, Application US/09608285A  
 ; Patent No. 6335013  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ford, John  
 ; APPLICANT: Mulero, Julio  
 ; APPLICANT: Yeung, George  
 ; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO CD39-LIKE  
 ; FILE REFERENCE: 28110/36570  
 ; CURRENT APPLICATION NUMBER: US/09/608,285A  
 ; CURRENT FILING DATE: 2000-06-30  
 ; PRIOR APPLICATION NUMBER: 09/583,231  
 ; PRIOR FILING DATE: 2000-05-26  
 ; PRIOR APPLICATION NUMBER: 09/557,800  
 ; PRIOR FILING DATE: 2000-04-25  
 ; PRIOR APPLICATION NUMBER: 09/481,238  
 ; PRIOR FILING DATE: 2000-01-11  
 ; PRIOR APPLICATION NUMBER: 09/370,265  
 ; PRIOR FILING DATE: 1999-08-09  
 ; PRIOR APPLICATION NUMBER: PCT/US99/16180  
 ; PRIOR FILING DATE: 1999-07-16  
 ; PRIOR APPLICATION NUMBER: 09/350,836  
 ; PRIOR FILING DATE: 1999-07-09  
 ; PRIOR APPLICATION NUMBER: 09/273,447  
 ; PRIOR FILING DATE: 1999-03-19  
 ; PRIOR APPLICATION NUMBER: 09/244,444  
 ; PRIOR FILING DATE: 1999-02-04  
 ; PRIOR APPLICATION NUMBER: 09/122,449  
 ; PRIOR FILING DATE: 1998-07-24  
 ; PRIOR APPLICATION NUMBER: 09/118,205  
 ; PRIOR FILING DATE: 1998-07-16  
 ; NUMBER OF SEQ ID NOS: 60  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 59  
 ; LENGTH: 15977  
 ; TYPE: DNA  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; OTHER INFORMATION: CD39-L4/L66 Gene Sequence  
 ; NAME/KEY: CDS  
 ; LOCATION: (245)..(461)  
 ; NAME/KEY: CDS  
 ; LOCATION: (1454)..(1533)  
 ; NAME/KEY: CDS  
 ; LOCATION: (2734)..(2877)  
 ; NAME/KEY: CDS  
 ; LOCATION: (4364)..(4439)  
 ; NAME/KEY: CDS  
 ; LOCATION: (4679)..(4714)  
 ; NAME/KEY: CDS  
 ; LOCATION: (5326)..(5414)  
 ; NAME/KEY: CDS  
 ; LOCATION: (5723)..(5802)  
 ; NAME/KEY: CDS  
 ; LOCATION: (6751)..(6812)  
 ; NAME/KEY: CDS  
 ; LOCATION: (7758)..(7859)



US-08-688-145-1

Query Match 9.2%; Score 259.2; DB 1; Length 3885;  
Best Local Similarity 73.4%; Pred. No. 4.7e-44;  
Matches 372; Conservative 0; Mismatches 129; Indels 6.

Qy	1666	AAACACAGATGATAAAGCTGCGAACCGTGGCACACGCTCATAGTCTTAGCTGCTGGGAG	1725
Db	1599	AAAAATTGAAGAAAGGGCCAGGTGTGATGGCTCATTTCCCTATAGTCCACGACCTTTGGGAG	1540
Qy	1726	GTTTGAGGAGGGAGGTGGCTTGAACACAGGTGTTCAAGGCCAGCCTGGGCCAACAATACAA	1785
Db	1539	GCCAAGGCGAGGAGGATCACTGAAGCCAGGAGTTCAGGACCAGCCTGGG--CTGTACCA	1483
Qy	1786	GATCCTGTCTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAAGAGAGAGGCGCGGGCGTGTGGC	1845
Db	1482	TGTTGAGACCCCTGTGTCAACTAAAAAACAAGAGAGGGGGGGCGGGCTGTGTAGC	1423
Qy	1846	TCACGCGCTGTAATCCAGACACTTTGGGAGGCCGAGCGCGGCGATCACCTGTGGTCAGGA	1905
Db	1422	TCACACCTGTATACCAAGCTGTGTGGAGGCCGAAGTGGGTGTATCTCTGAGGTCAGGA	1363
Qy	1906	GTTTGAACACAGCCTGGCCAAACATGGCAAAACCCCGTCTGTACTCAAAATGC--AAAAAT	1963
Db	1362	GTTTGTGACGAGCCTAGCCAAACATGGCGAAACCCCGTCTCCACTAAAAATACAAAAAAT	1303
Qy	1964	TAGCCAGGCGTGGTAGCAGGCACCTGTATATCCAGCTACTTTGGGAGGCTGAGCGAGAGA	2023
Db	1302	TAGCTGGGAGTGGTGGCGCGTGCCTGCAGTCCCGAGCTGCTTTGGGAGGCTTGAGCGAGAGA	1243
Qy	2024	ATCGCTTGAACCCAGGAGGTGGAGGTTGACGTAAGCTGAGATCGTGCGCTTGCACCTCCAG	2083
Db	1242	ATCGCTTGAACCCAGGAGCGAGGTTGCGAGTGAGCCAAAGATCGTGCCACTGCACCCAG	1183
Qy	2084	CCTGGCGCAAGAGCAAGACTCTGTCTCAGAAAAAATAAAAAAAAAAGAGAGAGAGAG	2143
Db	1182	CCTGGGCAAC-AGAGCAACCCCTATCTNAAAAAAAAAACAAGAAAGAAAGAGAGCAAAA	1124
Qy	2144	AAAGAGAACAAATATTTGGGAGAGAGG	2170
Db	1123	AAAAAGAAAATTAGCTGGATGTGGTGG	1097

RESULT 14  
US-09-735-934A-3  
; Sequence 3, Application US/09735934A

```

; Patent No. 6372468
;
; GENERAL INFORMATION:
;
; APPLICANT: LI, JIAYIN et al
;
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
;
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
;
; TITLE OF INVENTION: THEREOF
;
; FILE REFERENCE: CL000851
;
; CURRENT APPLICATION NUMBER: US/09/735,934A
;
; CURRENT FILING DATE: 2000-12-14
;
; NUMBER OF SEQ ID NOS: 4
;
; SOFTWARE: FastSeq for Windows Version 4.0
;
; SEQ ID NO. 3

```

```

; LENGTH: 43950
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-735-934A-3

```

	Query Match	Best Local Similarity	9.2%;	Score 258.6;	DB 4;	Length 43950;
	Matches 346;	Conservative	0;	Mismatches 89;	Indels	Gaps 3;
QY	1696	CACAGCTCATAGTTCTAGCTCTGGGAGGTTGAGCGGAGGATGCTTCAACACAGG	1755			
	35148	CTCAGCGCTGTAATCCACGACATTTGGAGGCCAAGGTGGGTGGATCACTTGTAGTCCAGG	35207			
QY	1756	TGTTCAAGCCCAAGCTGGGCAACATACAAGATCTGTCTCTCAAAAAA	1815			

Db	35208	AGTTTGAGACCGCCTGGTCAACATGGTGAACCCCTGTCTCTACTAAAAATAAAAA	35267
Qy	1816	AAAGAAAGAGAGAGCGCGGCGCTGTGTGCTCACGCCTGTAATCCACGACCTTGGGAGG	1875
Db	35268	AATTATCAGACGAACTGTGGCCCGGTGGCTCAGGCCTGTAATCCACGACCTTGGGAGG	35327
Qy	1876	CCGAGCCGGCGGATCACCTGTGGTCAGGAGTTTGAGACCGCCTGGCCAACTGGCAAA	1935
Db	35328	CCGAGGTGGGGAATCAC--GAGGTCAGAGTTTGAGACCGCCTGGCCAACTGGTGA	35385
Qy	1936	ACCCGCTCTGTACTCAAAAT-GCAAAAAATTAGCCAGCGGTGTGTAGCAGCACCTGTAAATC	1994
Db	35386	ACCCCATCTCTACTAAAAATAGAAAAAATTAGTGGGCATAGTGGCCAGCGCCTGTAATC	35445
Qy	1995	CCAGCTACTTGGGAGGCTGAGGCAGAGAAATCGCTTGAACCCAGAGAGGTGGAGTTGCAG	2054
Db	35446	CCAGCTACTTGGGAGGCTGAGGCAGAGAAATCGCTTGAACCCAGAGAGGTGGAGTTGCAG	35505
Qy	2055	TAAAGCTGAGATCGTGGCGTTGCACTCCAGCCTTGGCGACAAGACCAAGCTCTGTCTCAG	2114
Db	35506	TGAGCCGAGATCGTGCCACTGCNCTCCAGCCAGGGAC-AGAGTGAACNCTCCGCTCNA	35564
Qy	2115	AAAAAATAAAAAAAGAG	2133
Db	35565	AAAGAAAAAATAAATAG	35583

RESULT 15

```

US-09-798-096-10
; Sequence 10, Application US/09798096
; Patent No. 6399378
; GENERAL INFORMATION:
; APPLICANT: Donna T. Ward
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION
; FILE REFERENCE: RTS-0207
; CURRENT APPLICATION NUMBER: US/09798,096
; CURRENT FILING DATE: 2001-03-01
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 10
; LENGTH: 99500
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-798-096-10

```

	Query Match	9.1%	Score 257.6;	DB 4;	Length 99500;
	Best Local Similarity	69.1%;	Pred. No. 2.3e-43;		
	Matches 376;	Conservative 0;	Mismatches 149;	Indels 19;	Gaps 1;
QY	1616	TGTGTGATGCTTC	CCCAAGGTATATTTAACTATATATAGAGAGTTGTGTGACAAACAGAAT	1675	
Db	12586	TGTCCTAGTTTCTATAGCTGCTGTAATAAATACCATAAACTAGTGCCTTAAACAAC	12645		
QY	1676	GATAAAGTCGCAACCGTGGCACACGCTCATAGTTCTAGCTGCTTGGGAGTTGAGGAGG	1735		
Db	13646	ATAGAGCGCGGAACATTGGCTCATGCGCTGTAATCTCACAGTTTGGGAGGCCAAGGTGG	12705		
QY	1736	GAGGATGGCTTGAACACAGGTTGTTCAAGGCCACGCTGGGCAACATAACAAGATCCTTGTC	1795		
Db	12706	GAGTATCGCGGGAGCCCAAAATTCGAGACCAGTCTGGGCAACACGGTCAGTCCCTGTCT	12765		

QY	1796	CTCAAAAAA	AAAAAAAAA	AGAGAGAG	GGCGGCG	CTGTGTGGCT	CACGCCTGT	1855
Db	12766	CTGTTTAA	AAATAACA	CAATAACA	-----	-----	-----	12806
QY	1856	AATCCAGC	ACTTTGGG	AGGCGG	AGCGCGG	GATCACTCT	GTGTCAGG	1915
Db	12807	AATCCAGC	ACTTTGAG	AGGCGG	AGCGGCG	GATCACTCT	GTGTCAGG	12866
QY	1916	AGCCTGGC	CAACATGG	CAAAAC	CCCCCTG	TCTACTCAA	AATGCAAAA	1975
Db	12867	AGCCTGGC	CAACGTCG	CGAACTCC	ACTCTACT	AAAAATCA	AAAAATTC	12926







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:06:39 ; Search time 378.09 Seconds  
(without alignments)  
10937.110 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 2818  
Sequence: 1 adagaccatcattatga.....aaaaaaaaaaaaaaaaaa 2818

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA.\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US03\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2818	100.0	2818	9	US-09-966-880A-7
2	2174.6	77.2	6564	9	Sequence 7, Appli
3	2174.6	77.2	11204	9	Sequence 10, Appl
4	2172	77.1	2172	9	Sequence 35, Appl
5	603.4	21.4	2440	9	Sequence 15, Appl
6	295.2	10.5	13862	9	Sequence 1, Appli
7	295.2	10.5	13862	9	Sequence 5477, Ap
8	295.2	10.5	13862	9	Sequence 10204, A
9	292.6	10.4	7809	9	Sequence 1003, Ap
10	284.4	10.1	32249	9	Sequence 6094, Ap
11	281.4	10.0	65608	9	Sequence 7364, Ap
12	281.4	10.0	65608	10	Sequence 180, App
13	281.4	10.0	65608	10	Sequence 292, App
14	280.2	9.9	11821	10	Sequence 119, App
15	280	9.9	15297	9	Sequence 2857, Ap
16	278.2	9.9	32204	9	Sequence 3, Appli
17	275.4	9.8	3966	9	Sequence 8537, Ap
18	275.4	9.8	3966	9	Sequence 227, App
19	274.8	9.8	15362	10	Sequence 227, App
					Sequence 2856, Ap

C	20	274.2	9.7	11811	9	US-09-728-552-7	Sequence 7, Appli
	21	274	9.7	4963	10	US-09-764-877-2903	Sequence 2903, Ap
	22	273.6	9.7	174424	10	US-09-967-768A-314	Sequence 314, App
	23	273	9.7	26657	10	US-09-810-673A-3	Sequence 3, Appli
	24	271	9.6	271	9	US-09-966-880A-13	Sequence 13, Appli
	25	270.2	9.6	17286	10	US-09-764-877-3234	Sequence 3234, Ap
	26	269.8	9.6	43950	12	US-10-060-332-3	Sequence 3, Appli
	27	269.4	9.6	16086	10	US-09-764-877-2385	Sequence 2385, Ap
C	28	268	9.5	11677	9	US-10-082-830-134	Sequence 134, App
	29	267.8	9.5	35641	10	US-09-962-436-306	Sequence 306, App
	30	267.8	9.5	35641	10	US-09-880-107-2225	Sequence 2225, Ap
	31	266	9.4	56737	10	US-09-782-378A-17	Sequence 17, Appli
	32	266	9.4	110096	10	US-09-880-107-1542	Sequence 1542, Ap
C	33	265.6	9.4	9365	9	US-10-092-063-8	Sequence 8, Appli
	34	265.6	9.4	9365	12	US-10-091-085-8	Sequence 8, Appli
C	35	265	9.4	9903	9	US-09-764-891-8188	Sequence 8188, Ap
	36	264.8	9.4	32184	9	US-09-764-891-8538	Sequence 8538, Ap
	37	264.8	9.4	57130	10	US-09-835-081-3	Sequence 3, Appli
	38	264.4	9.4	145831	10	US-09-969-708-79	Sequence 79, Appli
C	39	264.4	9.4	145831	10	US-09-954-456-2116	Sequence 2116, Ap
	40	264	9.4	12452	10	US-09-764-877-3989	Sequence 3989, Ap
	41	263.6	9.4	4767	9	US-09-764-891-5474	Sequence 5474, Ap
C	42	263	9.3	10838	10	US-09-764-877-2698	Sequence 2698, Ap
	43	263	9.3	13608	10	US-09-861-846-3	Sequence 3, Appli
C	44	263	9.3	38374	10	US-09-880-107-3463	Sequence 3463, Ap
	45	262.8	9.3	3740	9	US-09-764-891-9984	Sequence 9984, Ap

ALIGNMENTS

RESULT 1  
US-09-966-880A-7  
; Sequence 7, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasaku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966, 880A  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 2818  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (80)...(673)  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: (1)...(79)  
; FEATURE:  
; NAME/KEY: 3'UTR  
; LOCATION: (677)...(2818)  
US-09-966-880A-7

Query Match 100.0%; Score 2818; DB 9; Length 2818;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2818; Conservative 0; Mismatches 0; Gaps 0;  
QY 1 AGAGAACCATCATTATGAAGTGAGACTTTTCTGGCCTGAGACTTTCGAGGAGGCAAGA 60  
|||||



QY 2221 GCCAATAAGGGATCCCTATTTGTCCTCTTTGGTGCTATTTGTCCTCCTAACAACTGCTCTT 2280  
Db |||||  
QY 2221 GCCAATAAGGGATCCCTATTTGTCCTCTTTGGTGCTATTTGTCCTCCTAACAACTGCTCTT 2280  
Db |||||  
QY 2281 GACAGTGAGAAAAATATTCAGAAATAACCATATCCCTGTGCGCGTTATTTACCTAGCAACCGT 2340  
Db |||||  
QY 2281 GACAGTGAGAAAAATATTCAGAAATAACCATATCCCTGTGCGCGTTATTTACCTAGCAACCGT 2340  
Db |||||  
QY 2341 TGCAATGAAGATGAGCAGATCCACAGGAAAACTTGAATGCACAACTGCTCTTATTTTAATC 2400  
Db |||||  
QY 2341 TGCAATGAAGATGAGCAGATCCACAGGAAAACTTGAATGCACAACTGCTCTTATTTTAATC 2400  
Db |||||  
QY 2401 TTATTTGACATAAGTTTGAAGAGTTAAAAATTTGTTACTTTCATGTATTCATTTATATTT 2460  
Db |||||  
QY 2401 TTATTTGACATAAGTTTGAAGAGTTAAAAATTTGTTACTTTCATGTATTCATTTATATTT 2460  
Db |||||  
QY 2461 TTATATTTTTCGCTCAATGATTTTATTAACATGATTTCCCTTTCTGATATATTTGA 2520  
Db |||||  
QY 2461 TTATATTTTTCGCTCAATGATTTTATTAACATGATTTCCCTTTCTGATATATTTGA 2520  
Db |||||  
QY 2521 AATGAGTCTCAAGCTTCATAAAATTTATAACTTTTAGAAATGATTTCTTAATAACAACGTAT 2580  
Db |||||  
QY 2521 AATGAGTCTCAAGCTTCATAAAATTTATAACTTTAGAAATGATTTCTTAATAACAACGTAT 2580  
Db |||||  
QY 2581 GPTAATTGAACATTCAGTAATGCTGCTACGAGCCATTTCTTTGATTTTGTAGTAACT 2640  
Db |||||  
QY 2581 GPTAATTGAACATTCAGTAATGCTGCTACGAGCCATTTCTTTGATTTTGTAGTAACT 2640  
Db |||||  
QY 2641 TTTATGACAGCAAAATTTGCTTCTGCTCACTTTTCAATCAGTTTAAATGAATAATAAT 2700  
Db |||||  
QY 2641 TTTATGACAGCAAAATTTGCTTCTGCTCACTTTTCAATCAGTTTAAATGAATAATAAT 2700  
Db |||||  
QY 2701 TTTGGAAGTGTGAAGATAAAATACCAATAAAATAATAATAAAAGTGATTTATATGAAGT 2760  
Db |||||  
QY 2701 TTTGGAAGTGTGAAGATAAAATACCAATAAAATAATAATAAAAGTGATTTATATGAAGT 2760  
Db |||||  
QY 2761 TAAATAAAAAATCAGTATGATGGAATAAACTTTGAAAAAATAAAATGAAAAA 2818  
Db |||||  
QY 2761 TAAATAAAAAATCAGTATGATGGAATAAACTTTGAAAAAATAAAATGAAAAA 2818  
Db |||||

RESULT 2

US-09-966-880A-10  
; Sequence 10, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966,880A  
; CURRENT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 10  
; LENGTH: 6564  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-966-880A-10

Query Match 77.2%; Score 2174.6; DB 9; Length 6564;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 2177; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
QY 617 CTTTGGCCCTGTATGAGTTTGATGACTTTACGAGCGCATTTTCGTAATTTGGGACTTTTGA 676

Db |||||  
3735 CTACAGCCCTGTATGAGTTTGATGACTTTACGAGAGCCATTTTCGTACTTTGGACATTTGA 3794  
QY |||||  
677 TAGCAACTTCCAGGAATGTCACACAGCATGAATAATCTCTGCTGAAGACAGTGGATAAAA 736  
Db |||||  
3795 TAGCAACTTCCAGGAATGTCACACAGCATGAATAATCTCTGCTGAAGACAGTGGATAAAA 3854  
QY |||||  
737 AACAGTCCCTCAAGTCTTCTCTGTTTTTATTTCAACTCTCACTTTCTTTAGAGTTTACA 796  
Db |||||  
3855 AACAGTCCCTCAAGTCTTCTCTGTTTTTATTTCAACTCTCACTTTCTTTAGAGTTTACA 3914  
QY |||||  
797 GAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTTGAAAAATAGAGAAGAAC 856  
Db |||||  
3915 GAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTTGAAAAATAGAGAAGAAC 3974  
QY |||||  
857 ACAGGTCTGCGCAGGAGCGTGTGCAATTTGGTGCAGTTTGAATGCAACATTTCCCCCTA 916  
Db |||||  
3975 ACAGGTCTGCGCAGGAGCGTGTGCAATTTGGTGCAGTTTGAATGCAACATTTCCCCCTA 4034  
QY |||||  
917 CTGGGAATAACAGAACTGCAGGACCTGGGAGGATCCTAAAGTGTCAACGTTTTTCTATGA 976  
Db |||||  
4035 CTGGGAATAACAGAACTGCAGGACCTGGGAGGATCCTAAAGTGTCAACGTTTTTCTATGA 4094  
QY |||||  
977 CTTTGTAGTATGAGCAGAGGATAGATCCTAAAAAGCATGGTGAGAGGATCAAAATG 1036  
Db |||||  
4095 CTTTGTAGTATGAGCAGAGGATAGATCCTAAAAAGCATGGTGAGAGGATCAAAATG 4154  
QY |||||  
1037 TTTTATATACATCCCTTTATTTGATTTGATTTGAGTTTAACAGTGGTGTAGTGATA 1096  
Db |||||  
4155 TTTTATATACATCCCTTTATTTGATTTGATTTGAGTTTAACAGTGGTGTAGTGATA 4214  
QY |||||  
1097 GATTTTCTATTTCTTCCCTTGACGTTTCAAGTAAACAACAACTCTTCCATCAGG 1156  
Db |||||  
4215 GATTTTCTATTTCTTCCCTTGACGTTTCAAGTAAACAACAACTCTTCCATCAGG 4274  
QY |||||  
1157 CCATGATCTATAGGACCTCCTAATGAGATATCTGGGTGATTTGACCCCAACACATCTC 1216  
Db |||||  
4275 CCATGATCTATAGGACCTCCTAATGAGATATCTGGGTGATTTGACCCCAACACATCTC 4334  
QY |||||  
1217 TCCAAAGCATTAATATCCCAATCATGCGCTGTATGTTTTTAATCAGCAGAGCATGTTTTA 1276  
Db |||||  
4335 TCCAAAGCATTAATATCCCAATCATGCGCTGTATGTTTTTAATCAGCAGAGCATGTTTTA 4394  
QY |||||  
1277 TGTGTTGTACAAAAGAGATTTGTTATGGTGGGATGAGGTATAGACCATGATGTCAC 1336  
Db |||||  
4395 TGTGTTGTACAAAAGAGATTTGTTATGGTGGGATGAGGTATAGACCATGATGTCAC 4454  
QY |||||  
1337 CTTCAAGCTACTTTAATAAAGGATCTTAAAAATGGCAGGAGGACTGTGAACAAGACACCC 4514  
Db |||||  
1397 TAATAATGGGTTGATGCTGGAAGTAGCAAAATCTTCTGGAAACCCCAACTCTTTTAAGAA 1456  
QY |||||  
4515 TAATAATGGGTTGATGCTGGAAGTAGCAAAATCTTCTGGAAACCCCAACTCTTTTAAGAA 4574  
Db |||||  
1457 GTCCCTAAATTTAGAAACACCCACAACTTTCACATATCATATAATAGCAAAATTTGAAGG 1516  
QY |||||  
4575 GTCCCTAAATTTAGAAACACCCACAACTTTCACATATCATATAATAGCAAAATTTGAAGG 4634  
Db |||||  
1517 AAGTTGCTTTGAATTTGGGAGAGGAAAAATCTATTTGGCTCTCTGTTGGTCTCTTCATCTCA 1576  
QY |||||  
4635 AAGTTGCTTTGAATTTGGGAGAGGAAAAATCTATTTGGCTCTCTGTTGGTCTCTTCATCTCA 4694  
Db |||||  
1577 GAAATGCCAATCAGGTCAAGGTTTGTACATTTTGTATGTTGTTGATGCTTCTCCCAAG 1636  
QY |||||  
4695 GAAATGCCAATCAGGTCAAGGTTTGTACATTTTGTATGTTGTTGATGCTTCTCCCAAG 4754  
Db |||||  
1637 GTATATTAATATATAAGAGAGTTGTGACAAAACAGATGATAAAGCTGCGAACCGTGGC 1696  
QY |||||  
4755 GTATATTAATATATAAGAGAGTTGTGACAAAACAGATGATAAAGCTGCGAACCGTGGC 4814  
Db |||||  
1697 ACACGCTCATAGTTCTAGCTGCTTGGGAGGTTTGGAGGAGGAGGATGCTTTGAACACAGGT 1756  
QY |||||

```

Db 4815 ACAGGCTCATTAGTTCCTAGCTGCTGGAGGTTGAGGAGGAGGATGCCTTGAACAGGT 4874
Qy 1757 GTTCAAGGCGCAGCTGGGCAACATAACAAGATCCTGCTCTCAAAAAAATAAAAAA 1816
Db 4875 GTTCAAGGCGCAGCTGGGCAACATAACAAGATCCTGCTCTCAAAAAAATAAAAAA 4934
Qy 1817 AAGAAAGAGAGAGGCGGCGGTGGTGGTGCACGCTGTAAATCCAGCACATTTGGGAGGC 1876
Db 4935 AAGAAAGAGAGAGGCGGCGGTGGTGGTGCACGCTGTAAATCCAGCACATTTGGGAGGC 4994
Qy 1877 CGAGCGGCGGATACCTGTGCTCAGGAGTTTGACACAGCCTGGCCAAATGCAAAA 1936
Db 4995 CGAGCGGCGGATACCTGTGCTCAGGAGTTTGACACAGCCTGGCCAAATGCAAAA 5054
Qy 1937 CCCCGTCTGTACTCAAAATGCAAAATAGCCAGCGCTGGTGAAGCAGCCTGTAATCCC 5114
Db 5055 CCCCGTCTGTACTCAAAATGCAAAATAGCCAGCGCTGGTGAAGCAGCCTGTAATCCC 5114
Qy 1997 AGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGAACCCAGGAGTGGAGTTCAGTA 2056
Db 5115 AGCTACTTGGGAGGCTGAGGAGGAGATCGCTTGAACCCAGGAGTGGAGTTCAGTA 5174
Qy 2057 AGCTGAGATCGTCCGCTGCACTCCAGCTGGCGGACAGAGACTCTGTCTCAGAA 2116
Db 5175 AGCTGAGATCGTCCGCTGCACTCCAGCTGGCGGACAGAGACTCTGTCTCAGAA 5234
Qy 2117 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2176
Db 5235 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 5294
Qy 2177 AAGCATTGCAAGGAAATGCTTTATCCACAAATGTAAGGAGCCAAATGAGGATCCC 2236
Db 5295 AAGCATTGCAAGGAAATGCTTTATCCAAATAATGTAAGGAGCCAAATGAGGATCCC 5354
Qy 2237 TATTGTCTCTTTGGTGTCTATTGTGCTTAACTGCTTTGACAGTGAAGAGAGAG 2296
Db 5355 TATTGTCTCTTTGGTGTCTATTGTGCTTAACTGCTTTGACAGTGAAGAGAGAG 5414
Qy 2297 TTCAGATAACCATATCCCTGCGCTTATACCTACCAACCTTGAATGAGATGAGC 2356
Db 5415 TTCAGATAACCATATCCCTGCGCTTATACCTACCAACCTTGAATGAGATGAGC 5474
Qy 2357 AGATCCAGAGAAATGTAATGACACACTGCTTTATTTAACTTATTTGATGATGAT 2416
Db 5475 AGATCCAGAGAAATGTAATGACACACTGCTTTATTTAACTTATTTGATGATGAT 5534
Qy 2417 TGTAAAGAGTTAAATTTGCTTACTTCTATGATGATGATGATGATGATGATGATG 2476
Db 5535 TGTAAAGAGTTAAATTTGCTTACTTCTATGATGATGATGATGATGATGATGATG 5594
Qy 2477 CTAAATGATTTTATTAACATGATTTCTTTCTGATATATTCAAATGGAGTCTCAAGC 2536
Db 5595 CTAAATGATTTTATTAACATGATTTCTTTCTGATATATTCAAATGGAGTCTCAAGC 5654
Qy 2537 TTCAATAATTTAATCTTGAATGATGATGATGATGATGATGATGATGATGATGATG 2596
Db 5655 TTCAATAATTTAATCTTGAATGATGATGATGATGATGATGATGATGATGATGATG 5714
Qy 2597 AGTAATGGTCTACGAGGCAATTTCTTTGATTTTATGATTTTATGACAGCAATTT 2656
Db 5715 AGTAATGGTCTACGAGGCAATTTCTTTGATTTTATGATTTTATGACAGCAATTT 5774
Qy 2657 TGCTTCTGGCTCCTTCAATCAGTTTAAATAAATGATGATGATGATGATGATGATG 2716
Db 5775 TGCTTCTGGCTCCTTCAATCAGTTTAAATAAATGATGATGATGATGATGATGATG 5834
Qy 2717 ATAAATACCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 2776
Db 5835 ATAAATACCAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 5894
Qy 2777 TATGATGGAATAAATGTA 2797
Db 5895 TATGATGGAATAAATGTA 5915

```

RESULT 3

```

US-09-966-880A-35
; Sequence 35, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 11204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-35

```

```

Query Match 77.2%; Score 2174.6; DB 9; Length 11204;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 2177; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 617 CTTTGGCCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTTCGTACTTTGGGACTTTTCA 676
Db 8950 CTACAGCCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTTCGTACTTTGGGACTTTCA 9009
Qy 677 TAGCAACTTCCAGGAATGTACACACGATGAATATCTCTGCTGAAGACAGTGGATGATAA 736
Db 9010 TAGCAACTTCCAGGAATGTACACACGATGAATATCTCTGCTGAAGACAGTGGATGATAA 9069
Qy 737 ACAGTCTTCAAGTCTTCTGTTTATTTTCAACTCTCACTTCTTAGAGTTTACA 796
Db 9070 AACAGTCTTCAAGTCTTCTGTTTATTTTCAACTCTCACTTCTTAGAGTTTACA 9129
Qy 797 GAAAAATATTTATATACGACTCTTTAAAGACATCTATGCTTGAATGCAACATTTGCCCTTA 916
Db 9130 GAAAAATATTTATATACGACTCTTTAAAGACATCTATGCTTGAATGCAACATTTGCCCTTA 9249
Qy 857 ACAGTCTTGGCCAGGAGGAGTGTGCAATTTGGTGCAGTTTGAATGCAACATTTGCCCTTA 916
Db 9190 ACAGTCTTGGCCAGGAGGAGTGTGCAATTTGGTGCAGTTTGAATGCAACATTTGCCCTTA 9249
Qy 917 CTGGGAATAACAGAACTGCAGACCTGGGAGCATCTTAAAGTGTCAACGTTTTCATGTA 976
Db 9250 CTGGGAATAACAGAACTGCAGACCTGGGAGCATCTTAAAGTGTCAACGTTTTCATGTA 9309
Qy 977 CTTTGTAGTAGGATGAGAGAGAGTGTGCAATTTGGTGCAGTTTGAATGCAACATTTGCCCTTA 1036
Db 9310 CTTTGTAGTAGGATGAGAGAGAGTGTGCAATTTGGTGCAGTTTGAATGCAACATTTGCCCTTA 9369
Qy 1037 TTTTATATCAACATCTTTTATTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 1096
Db 9370 TTTTATATCAACATCTTTTATTTATTTGATTTGATTTGATTTGATTTGATTTGATTTGATTT 9429
Qy 1097 GATTTTCTATTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1156
Db 9430 GATTTTCTATTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 9489
Qy 1157 CCATGATCTATAGGACCTCTTAATGAGATGATCTGGGTGATTTGATGATGATGATGATGATG 1216
Db 9490 CCATGATCTATAGGACCTCTTAATGAGATGATCTGGGTGATTTGATGATGATGATGATGATG 9549

```

QY	1217	TCCAAAGCAATTAATATCCAAATCATGCGCTGTATGTTTTTAATCAGCAGAAAGCATGTTTTTA	1276
Db	9550	TCCAAAGCAATTAATATCCAAATCATGCGCTGTATGTTTTTAATCAGCAGAAAGCATGTTTTTA	9609
QY	1277	TCCTTTGTACAAAGAGATTGTTATGGGTGGGATGGAGGTATAGACCATGCATGTCAC	1336
Db	9610	TCCTTTGTACAAAGAGATTGTTATGGGTGGGATGGAGGTATAGACCATGCATGTCAC	9669
QY	1337	CTTCAAGCTACTTTTAATAAAGGATCTTAAATGGGCAGGAGGACTGTGAACACAGACACCC	1396
Db	9670	CTTCAAGCTACTTTTAATAAAGGATCTTAAATGGGCAGGAGGACTGTGAACACAGACACCC	9729
QY	1397	TAATAATGGGTGTATGTCGAAGTAGCAAAATCTCTGAAACGCAAACTCTTTTAAGGAA	1456
Db	9730	TAATAATGGGTGTATGTCGAAGTAGCAAAATCTCTGAAACGCAAACTCTTTTAAGGAA	9789
QY	1457	GTCCCTAATTTAGAAACACCCCAAACTTCACATATCATATAATTAGCAAACTTGAAGG	1516
Db	9790	GTCCCTAATTTAGAAACACCCCAAACTTCACATATCATATAATTAGCAAACTTGAAGG	9849
QY	1517	AAGTTGCTTGAATGTTGGGAGAGGAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA	1576
Db	9850	AAGTTGCTTGAATGTTGGGAGAGGAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA	9909
QY	1577	GAATGCCAATCAGGTCAAAGTTTGCTACATTTTGCTATGCTGTATGCTTCTCCAAAG	1636
Db	9910	GAATGCCAATCAGGTCAAAGTTTGCTACATTTTGCTATGCTGTATGCTTCTCCAAAG	9969
QY	1637	GTATATTACTATATAAGAGAGTTGTGACAAACAGAAATGATAAGCTCGGAACCGTGGC	1696
Db	9970	GTATATTACTATATAAGAGAGTTGTGACAAACAGAAATGATAAGCTCGGAACCGTGGC	10029
QY	1697	ACACGCTCATAGTCTTAGCTGCTGGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGT	1756
Db	10030	ACACGCTCATAGTCTTAGCTGCTGGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGT	10089
QY	1757	GTTCAAGGCCAGCTGGGCAACATACAAAGATCCTGTCTCTCAAAAAAATAAAAAA	1816
Db	10090	GTTCAAGGCCAGCTGGGCAACATACAAAGATCCTGTCTCTCAAAAAAATAAAAAA	10149
QY	1817	AGAAGAGAGAGGGCCGGCGTGTGGCTCAGCGCTGTAAATCCAGCAGCTTTGGGAGGC	1876
Db	10150	AGAAGAGAGAGGGCCGGCGTGTGGCTCAGCGCTGTAAATCCAGCAGCTTTGGGAGGC	10209
QY	1877	CGAGCGGGCGGATCACCTGTGGTCAAGGATTTGACACAGCCTGGCCCAACATGGCAAAA	1936
Db	10210	CGAGCGGGCGGATCACCTGTGGTCAAGGATTTGACACAGCCTGGCCCAACATGGCAAAA	10269
QY	1937	CCCCGTCTGTACTCAAAATGCAAAATTAGCCAGGCGGTGGTAGCAGGCACCTGTAATCCC	1996
Db	10270	CCCCGTCTGTACTCAAAATGCAAAATTAGCCAGGCGGTGGTAGCAGGCACCTGTAATCCC	10329
QY	1997	AGCTACTTGGGAGGCTGAGGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGTTGAGTA	2056
Db	10330	AGCTACTTGGGAGGCTGAGGCAGGAGAAATCGCTTGAACCCAGGAGGTGGAGTTGAGTA	10389
QY	2057	AGCTGAGATCGTGGCGTTCGACTCCAGCTGGCGCAAGAGCAAGCACTGTGCTCAGAA	2116
Db	10390	AGCTGAGATCGTGGCGTTCGACTCCAGCTGGCGCAAGAGCAAGCACTGTGCTCAGAA	10449
QY	2117	AAAAAATAAAG	2176
Db	10450	AAAAAATAAAG	10509
QY	2177	AGCATTCGAAGAAATGTGCTTTATCCAAACAAATGTAAAGGCAATAGGGATCCC	2236
Db	10510	AGCATTCGAAGAAATGTGCTTTATCCAAACAAATGTAAAGGCAATAGGGATCCC	10569
QY	2237	TATTTGCTCTTTTGGTGCTATTTGTCCCTTAACACTGTCTTTTGACAGTGAGAAAAATA	2296
Db	10570	TATTTGCTCTTTTGGTGCTATTTGTCCCTTAACACTGTCTTTTGACAGTGAGAAAAATA	10629
QY	2297	TTCAGAATAACCATATCCCTGTGCGGCTTATTACCTAGCAACCCCTTGCAATGAAGATCAGC	2356

Db	10630	TTCAGATAACCATATCCCTGTGCGGTTATTACCTAGCAACCCCTTGCAATGAAGATGAGC	10689
QY	2357	AGATCCACAGGAAAACTTGAATGCACAACCTGCTCTTATTTAACTCTTATTTGACATAAGTT	2416
Db	10690	AGATCCACAGGAAAACTTGAATGCACAACCTGCTCTTATTTAACTCTTATTTGACATAAGTT	10749
QY	2417	TGTAAGAAGAGTTAAAAATTTACTTCTCATGTATTTCATTTATATTTATTTTTCGCT	2476
Db	10750	TGTAAGAAGAGTTAAAAATTTACTTCTCATGTATTTCATTTATATTTATTTTTCGCT	10809
QY	2477	CTAATGATTTTTTATTAAACATGATTTCCCTTTCTGATATATTGAAATGGAGTCTCAAGC	2536
Db	10810	CTAATGATTTTTTATTAAACATGATTTCCCTTTCTGATATATTGAAATGGAGTCTCAAGC	10869
QY	2537	TTTCATAAATTTATAAATTTAGAAATGATTTCTTAATAACAAGCTATGTAATTTGTAACATTGC	2596
Db	10870	TTTCATAAATTTATAAATTTAGAAATGATTTCTTAATAACAAGCTATGTAATTTGTAACATTGC	10929
QY	2597	AGTAATGGTCTACGAGCCATTCTCTTGATTTTTAGTAAACTTTTATGACAGCAAAAT	2656
Db	10930	AGTAATGGTCTACGAGCCATTCTCTTGATTTTTAGTAAACTTTTATGACAGCAAAAT	10989
QY	2657	TGCTTCTGGCTCACTTTCAATCAGTTAAATAAATGATAATAATTTTGGAGCTGTGAAG	2716
Db	10990	TGCTTCTGGCTCACTTTCAATCAGTTAAATAAATGATAATAATTTTGGAGCTGTGAAG	11049
QY	2717	ATAAATACCAATAAATAAATAAATAAAGTGATTTATATGACGTTAAATAAATAAATCAG	2776
Db	11050	ATAAATACCAATAAATAAATAAATAAAGTGATTTATATGACGTTAAATAAATAAATCAG	11109
QY	2777	TATGATGGAATAAATTTGAAA 2797	
Db	11110	TATGATGGAATAAATTTGAGA 11130	
RESULT 4			
US-09-966-880A-15			
; Sequence 15, Application US/09966880A			
; Patent No. US20020164743A1			
; GENERAL INFORMATION:			
; APPLICANT: Honjo, Tasuku			
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE			
; FILE REFERENCE: 06501-088001			
; CURRENT APPLICATION NUMBER: US/09/966, 880A			
; CURRENT FILING DATE: 2001-09-28			
; PRIOR APPLICATION NUMBER: PCT/JP00/01918			
; PRIOR FILING DATE: 2000-03-28			
; PRIOR APPLICATION NUMBER: JP 11-371382			
; PRIOR FILING DATE: 1999-12-27			
; PRIOR APPLICATION NUMBER: JP 11-178999			
; PRIOR FILING DATE: 1999-06-24			
; PRIOR APPLICATION NUMBER: JP 11-87192			
; PRIOR FILING DATE: 1999-03-29			
; NUMBER OF SEQ ID NOS: 36			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 15			
; LENGTH: 2172			
; TYPE: DNA			
; ORGANISM: Homo sapiens			
US-09-966-880A-15			
Query Match 77.1%; Score 2172; DB 9; Length 2172;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2172; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	623	CCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTCCGTACTTTGGGACTTTGTATAGCAA	682
Db	1	CCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTCCGTACTTTGGGACTTTGTATAGCAA	60
QY	683	CTTCCAGGAATGTCACACGATGAATATCTCTGCTGAAGACAGTGGATATAAATAACAGT	742





Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966, 880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
PRIOR APPLICATION NUMBER: JP 11-371382  
PRIOR FILING DATE: 1999-12-27  
PRIOR APPLICATION NUMBER: JP 11-178999  
PRIOR FILING DATE: 1999-06-24  
PRIOR APPLICATION NUMBER: JP 11-87192  
PRIOR FILING DATE: 1999-03-29  
NUMBER OF SEQ ID NOS: 36  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 2440  
TYPE: DNA  
ORGANISM: Mus musculus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (93)...(686)  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: (1)...(92)  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: (690)...(2440)  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)...(2440)  
OTHER INFORMATION: n = A,T,C or G  
US-09-966-880a-1

Query Match 21.4%; Score 603.4; DB 9; Length 2440;  
Best Local Similarity 69.4%; Pred. No. 2e-120;  
Matches 1006; Conservative 0; Mismatches 351; Indels 92; Gaps 10;

QY	4	GAACCATCATTAATGAAGTGTAGATTTTCTGGCCCTGAGACTTTCGAGAGGAGCAAG-AAG	62
Db	16	GAAGCAGCTTGTCTTGAAGCAGCTTCTTGGCCCTTAAGACTTTGAGGAGTCAAGAAG	75
QY	63	ACACTCTGACACCATATGGACGCTCTTGTATGAACGGAGGAAGTTCTTTACCAAT	122
Db	76	TACGCTGGAGACCGATATGGACAGCCTTCTGTATGAAGCAAAAGAGTTCTTTACCAAT	135
QY	123	TCAAAATATCGCGTGGGCTAAGGTCGGCGTGAGACCTTACCTGTGCTAGCTAGTGAAGA	182
Db	136	TCAAAATATCGCGTGGGCAAGGACGCGATGAGACCTTACCTGTGCTAGCTAGTGAAGA	195
QY	183	GGCGTGACAGTGTACATCCTTTTCTACTGAGCTTTGGTTATCTTCGCAATAAGACGGCT	242
Db	196	GGAGAGATATGCCACCTCTGCTCACTGGACTTCGGCCACCTTCGCAACAAGTCTGGCT	255
QY	243	GCCAGTGAATGTCTTCTCCGCTACATCTCGGACTTGGGACCTAGACCTTGGCGGCT	302
Db	256	GCCAGTGAATGTGTCTTCTCCGCTACATCTCGGACTTGGGACCTAGACCTTGGCGGCT	315
QY	303	GCTACGCTACCTCGTTTCACTCTCTGGAGCCCTTGTCTAGCTGTGCGCGACATGTGG	362
Db	316	GTTACGCGTACCTCGTTTCACTCTCTGGAGCCCTTGTCTAGCTGTGCGCGACATGTGG	375
QY	363	CCGACTTTCTGGAGGGAACCCCACTCAGTCTGAGGATCTTTCACCGCGGCTCTACT	422
Db	376	CTGAGTTCTGAGTGGAAACCTTAACCTCAGCTGAGGATTTTTCACCGCGGCTCTACT	435
QY	423	TCTGTGAGACCGCAGGCTGAGCCCGAGGGCTGGCGGCTGCGACCGCGCGGGTGC	482
Db	436	TCTGTGAGACCGCAGGCTGAGCCCGAGGGCTGGCGGCTGCGACCGCGCGGGTGC	495

RESULT 6

US-09-764-891-5477

; Sequence 5477, Application US/09764891

; Publication No. US2003007780A1

QY	483	AAATAGCATCATGACCTTCAAGATATTTTTACTGCTGAATACATTTTGTAGAAAACC	542
Db	496	AGATCGGATCATGACCTTCAAGACTATTTTACTGCTGAATACATTTGTAGAAAATC	555
QY	543	ATGAAGAAGCTTTCAAGCCCTGGGAAGGCTGCATGAATAATTCAGTTCGTCTCTCCAGAC	602
Db	556	GTGAAGAAGCTTTCAAGCCCTGGGAAGGCTGCATGAATAATTCGTCTCCGCTTAACACGAC	615
QY	603	AGCTTCGGCGCATCTTTTCCCTGTATGAGGTGATGACTTACGAGACCATTTCTGTA	662
Db	616	AACCTCGGCGCATCTTTTCCCTGTATGAGGTGATGACTTACGAGACCATTTCTGTA	675
QY	663	CTTTGGGACTTTGATAGCAACTTCCAGGAATGTCAACACGATGAAATATCTCTGCTGAA	722
Db	676	TGTTGGGATTTTGAAGCAACCTTCTCGAATGTCAACGATGAAATTTCT---CTGAA	732
QY	723	GACAGTGGATAAAAACAGCTCCCTCAAGTCTTCTCTGTTTTTATTTTCACTCTCACTT	782
Db	733	GAGACTGGATAGAAAACAAACCCCTTCA--ACTACATGTTTTTCTTCTTAAGTACTCACT	790
QY	783	TCTTAGAGTTTACAGAAAAAATATTTATATACGACTCTTTAAAAAGATCTATGCTTGAA	842
Db	791	TTATAAGTGTAGGGGAAA-----TTATATGACTTTTAAAAAATACTTGAGCTGCAC	843
QY	843	AATAGAGAAGGAACACAGGCTTGCCAGGACGCTGCTGCAATTTGGTGCAGTTTGAATGC	902
Db	844	AGGACCCAGAGCAATGATTAAC-----TCAGCTTCTGCTGTC	882
QY	903	AACATTGTCCTTACT--GGGAATAACAGAACTGCGAGGACCTGGGAGCATCTTAAGTGT	961
Db	883	AACATGCCATCTACTGGGAACAGCATAACTTCCAGACTTTGGGCTGCTGAATGATGCTC	942
QY	962	AACGTTTTTCTATGACTTTTAGTGTAGGATGAGACAGCAAGTACATCTTAAAGCATGG	1021
Db	943	TTTTTTTTCAACAGCAT-----GGAAAAGCATAT	971
QY	1022	TGAGAGGATCAAAATGTTTTTATATCAACATCCTTATTTATTTGATTGATTGAGTTAACA	1081
Db	972	GGAGACGACACAGCTTTGTTACACCCACCTGCTGTTCTTGTATTGATTGATTGATTCTCA	1031
QY	1082	GTGGTGTATGATAGATTTTCTTATTTCTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	1141
Db	1032	GGGTATCATGACGGATTTCTTATTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT	1084
QY	1142	AACCTTCCATCAGGCGCATGATCTATAGGACCTCTATGAGATGATCTGGGTGATTTGTG	1201
Db	1085	---GGTCTTTCTGCAAGGTCACGGGCTGCTTACAGTCTCTGCTGAGCAAT-----	1137
QY	1202	ACCCAAACCATCTCTCAAGCATTTAATATCCAAATCATGCGCTGTATGTTTAAATCAGC	1261
Db	1138	-CACAGCCATCTCTCAAAAACATTAATCTACAGGCACATGCTGTATGTTTCA-----	1191
QY	1262	AGAAGCATGTTTTATGTTTGTACAAAAGATGTTTATGGGTGGGATGGAGGTATAG	1321
Db	1192	-----CTGTCCGCTGCTGTTTTTACATTTGTATGTGAAGGGCTTGGGGTGGATTTGA	1245
QY	1322	ACCATGATGATCAGCTTCAAGCTTCTTAAATGAAGTACTTAAATGGGCGAGGAGT	1381
Db	1246	AGATGACCATGCTGCTTGGGTGATTCTAATAAGGATCTTAAATGCAATGAGGACT	1305
QY	1382	GTGAACAGACACCTTAATATGGTTGATGCTGTGAAGTACAAATCTTCTGGAAACGCA	1441
Db	1306	ACGAAGAATCACTCTGAAATGAGTTTCAACGCTTCAAGCAAAATCCCTTGGAAACACA	1365
QY	1442	AACCTTTTT	1450
Db	1366	GACTCTTTT	1374

```

; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5477
; LENGTH: 13862
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-891-5477

Query Match      10.5%; Score 295.2; DB 9; Length 13862;
Best Local Similarity 74.6%; Pred. No. 3e-53;
Matches 397; Conservative 0; Mismatches 133; Indels 2; Gaps 2;

QY 1667 AACAGAATGATAAGCTGCGAACCCTGGCACACGCTCAATAGTCTTAGCTTGGGAGG 1726
DB 1508 AAACCTTAATAAGCTGGCTCTGTGGCTCATGCTGTGATCCCGACACTTTGGGAGG 1567

QY 1727 TTGAGGAGGAGGATGGCTTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATACAAG 1786
DB 1568 CTGAGGAGGAGGATCACTTGAGGTGAGGAGTTTGAGACGAGCTGGCCAAACATGGCGAA 1627

QY 1787 ATCCTGTCTCTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1845
DB 1628 ACCCATCTCTACTATAAATACAAAAGTTAGCCAGGCATGGGCGGCGATGGTGGTGGC 1687

QY 1846 TCACGCTGTAAATCCAGCACTTTGGGAGCGGAGCGGCGGCGGATCACCTGTGGTCAGGA 1905
DB 1688 TCATGCTGTAAATCCAGCACTTTGGGAGCGGAGGTTGGCGGATCACCTGAGTCCAGGA 1747

QY 1906 GTTTGAGACCGCCTGGCCAAACATGCAAAACCCCTCTGTACTCAAAATGCAAAATTA 1965
DB 1748 GTTCAAGACCGCCTGGCAACATGCAAAACCCCTCTGTACTCAAAATGCAAAATTA 1807

QY 1966 GCCAGCGTGTAGCAGGACCTGTAATCCAGCTACTTTGGGAGGCTGAGGCGAGGAGAAAT 2025
DB 1808 GCCAGGCTTATGTAGTGGTGGCGGCTATATCCAGCTACTCGGAAGGCTGAGGCGAGGAGAAAT 1867

QY 2026 CGCTTGAACCCAGGAGGTGGAGTTGCAGTAAGCTGAGATCGTGGCTTGCATCCAGCC 2085
DB 1868 CGTTTGAAGCCAGGAGCGGAGTTTGCACCGAGCGGAGATTTGCCACTGCATCCAGCC 1927

QY 2086 TGGGCGACAGACGAGACTCTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 2145
DB 1928 TGGCGGATAA-AGCGAGACTCTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAA 1986

QY 2146 AGAAACAATATTTTGGGAGAGAGAGGATGGGAAGCATTTGCAAGGAAATTTGTG 2197
DB 1987 AGAAAAAATTAGCCAGCGTGGTGGCATGCACCTGTAGTCTAGTACTTTGGG 2038

RESULT 7
US-09-764-891-10204
; Sequence 10204, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10204
; LENGTH: 13862
; TYPE: DNA
; ORGANISM: Homo sapiens
```

```

US-09-764-891-10204
Query Match      10.5%; Score 295.2; DB 9; Length 13862;
Best Local Similarity 74.6%; Pred. No. 3e-53;
Matches 397; Conservative 0; Mismatches 133; Indels 2; Gaps 2;

QY 1667 AACAGAATGATAAGCTGCGAACCCTGGCACACGCTCAATAGTCTTAGCTTGGGAGG 1726
DB 1508 AAACCTTAATAAGCTGGCTCTGTGGCTCATGCTGTGATCCCGACACTTTGGGAGG 1567

QY 1727 TTGAGGAGGAGGATGGCTTTGAACACAGGTGTTCAAGGCCAGCCTGGGCAACATACAAG 1786
DB 1568 CTGAGGAGGAGGATCACTTGAGGTGAGGAGTTTGAGACGAGCTGGCCAAACATGGCGAA 1627

QY 1787 ATCCTGTCTCTCAAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1845
DB 1628 ACCCATCTCTACTATAAATACAAAAGTTAGCCAGGCATGGGCGGCGATGGTGGTGGC 1687

QY 1846 TCACGCTGTAAATCCAGCACTTTGGGAGCGGAGCGGCGGCGGATCACCTGTGGTCAGGA 1905
DB 1688 TCATGCTGTAAATCCAGCACTTTGGGAGCGGAGGTTGGCGGATCACCTGAGTCCAGGA 1747

QY 1906 GTTTGAGACCGCCTGGCCAAACATGCAAAACCCCTCTGTACTCAAAATGCAAAATTA 1965
DB 1748 GTTCAAGACCGCCTGGCAACATGCAAAACCCCTCTGTACTCAAAATGCAAAATTA 1807

QY 1966 GCCAGCGTGTAGCAGGACCTGTAATCCAGCTACTTTGGGAGGCTGAGGCGAGGAGAAAT 2025
DB 1808 GCCAGGCTTATGTAGTGGTGGCGGCTATATCCAGCTACTCGGAAGGCTGAGGCGAGGAGAAAT 1867

QY 2026 CGCTTGAACCCAGGAGGTGGAGTTGCAGTAAGCTGAGATCGTGGCTTGCATCCAGCC 2085
DB 1868 CGTTTGAAGCCAGGAGCGGAGTTTGCACCGAGCGGAGATTTGCCACTGCATCCAGCC 1927

QY 2086 TGGGCGACAGACGAGACTCTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 2145
DB 1928 TGGCGGATAA-AGCGAGACTCTCTCAGAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAA 1986

QY 2146 AGAAACAATATTTTGGGAGAGAGAGGATGGGAAGCATTTGCAAGGAAATTTGTG 2197
DB 1987 AGAAAAAATTAGCCAGCGTGGTGGCATGCACCTGTAGTCTAGTACTTTGGG 2038

RESULT 8
US-10-205-428-1003
; Sequence 1003, Application US/10205428
; Publication No. US20030108907A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PALL7C1
; CURRENT APPLICATION NUMBER: US/10/205,428
; CURRENT FILING DATE: 2002-07-26
; PRIOR APPLICATION NUMBER: 09/764,892
; PRIOR FILING DATE: 2001-01-17
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: 60/214,886
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 60/217,487
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,758
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/220,963
; PRIOR FILING DATE: 2000-07-26
; PRIOR APPLICATION NUMBER: 60/217,496
; PRIOR FILING DATE: 2000-07-11
; PRIOR APPLICATION NUMBER: 60/225,447
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/218,290
; PRIOR FILING DATE: 2000-07-14
```









; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 15297  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-003-295-3

```
Query Match          9.9%; Score 280; DB 9; Length 15297;
Best Local Similarity 74.9%; Pred. No. 6.3e-50;
Matches 370; Conservative 0; Mismatches 110; Indels 14; Gaps 1;

Qy 1641 ATTAACATATATAGAGAGTTGTGCACAAACAGAAATGATAAGCTGCGAACCCGTGGCACAC 1700
Db 11651 ACTCCCTTAATGCGCAACCTCCACCAGGAGGAGAAATAGAAATAACCTGGCCAGTTGCTCAC 11710

Qy 1701 GCTCATAGTTCTAGCTGCTTGGGAGGTTGAGGAGGAGGAGTGTGGAACACAGGTGTTTC 1760
Db 11711 GCCTGTATCCCGACACTTTGGGAGGCTGAGCTGGGTGGATCATTGAGCCCGAGGAGTTC 11770

Qy 1761 AAGGCCAGCCTGGGCAACATAACAAGATCCTGCTCTCAAAAAAATAAAAAAAGA 1820
Db 11771 AAGATCAGCTTGACACACAGTGAACCTCCATCTGTACAAAAATACAAAAATA----- 11825

Qy 1821 AAGAGAGAGGCGCGGCGTGGTGGCTCACGCCCTGTATATCCAGCAGCTTTGGGAGGCCGAG 1880
Db 11826 -----GACTGGGCACGCTGGCTCACACCTGTATATCCAGCAGCTTTGGGAGGCCGAG 11876

Qy 1881 CCGGGCGGATCACCTGTGGTCAGGAGTTTGAGACCCAGCCTGGCCACATGGCAAAACCCC 1940
Db 11877 GCAGGTGGATCACCTGTGGTCAGGAGTTTGAGACCCAGCCTGGCCACATGGTGAACCCCC 11936

Qy 1941 GTCTGTACTCAAAATGCAAAATTAGCCAGGCGTGGTAGCAGGCACCTGTAATCCAGCT 2000
Db 11937 ATCTTACTAAATAACAAAAATAGCCAGGCATGGTGGCAGCTGTGCTGTATCCAGCT 11996

Qy 2001 ACTTGGGAGGCTGAGGAGAGAGAAATCGCTTTGAACCCAGGAGGTGGAGGTTCAGTAAGCT 2060
Db 11997 ACTTGGGAGGCTGAGGTGGGAGAAATGCTTTGAACCCAGGAGGCGGAGGTGCAGTGAGCC 12056

Qy 2061 GAGATCGTGCCTTGGCAGCTCCAGCCTGGGCGACAGAGAGAGTCTGTCTCAGAAAAAA 2120
Db 12057 GAGATTGTGCCACTGCAGCTCCAGCCTGGGCGGACAGAGTGAAGTCCATCTCAAAAAAA 12116

Qy 2121 AAAAAAAGAGAGA 2134
Db 12117 CCAAAAAACAAAA 12130
```

Search completed: June 19, 2003, 11:33:58  
Job time : 381.09 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 14, 2003, 18:09:48 ; Search time 115.983 Seconds  
(without alignments)  
6475.074 Million cell updates/sec

Title: US-09-966-880A-7

Perfect score: 4948  
Sequence: 1 adagacaccatcattatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: BLOSUM62 Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1815940

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-O=Cpn2\_1/USPTO.spool/US09966880/runat\_14062003\_175523\_10297/app\_query.fasta\_1.9493  
-DB=A\_Geneseq\_101002 -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LIST=45 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-MODE=LOCAL -OUTFMT=ptc -THR\_SCORE=100 -THR\_MIN=0 -ALIGN=15  
-USER=US09966880 -ECN\_1\_1\_632 -runat\_14062003\_175523\_10297 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1086	21.9	198	21	AAB24198	Human activation-1
2	1008	20.4	198	21	AAB24197	Mouse activation-1
C 3	417	8.4	188	22	ABG11932	Novel human diagno
C 4	402	8.1	126	22	ABB11478	Human secreted pro
C 5	394	8.0	110	21	ABA43292	Human ORFX ORF3056
6	390	7.9	384	20	AA42383	Amino acid sequenc
7	390	7.9	384	22	AAB39075	Human secreted pro
8	390	7.9	384	23	AB55784	Human polypeptide
9	388	7.8	119	22	AA83429	Human immune/haema
10	388	7.8	384	21	AA84437	Amino acid sequenc
C 11	386	7.8	107	22	AU31902	Novel human secret
C 12	383.5	7.8	134	22	AB12093	Human secreted pro
C 13	381.5	7.7	239	22	AB90850	Human secreted pro
C 14	381.5	7.7	310	22	AB90647	Human secreted pro
C 15	378	7.7	165	21	ABA2917	Human ORFX ORF2681
C 16	377	7.6	381	22	AU30235	Novel human secret
C 17	376	7.6	175	22	AU31725	Novel human secret
C 18	375.5	7.6	124	23	AA80385	Human neurofibroma
C 19	375.5	7.6	382	22	AU31818	Novel human secret
C 20	375.5	7.6	382	22	AU32610	Novel human secret
C 21	375.5	7.6	382	22	AU32707	Novel human secret
C 22	371.5	7.5	121	22	AU32891	Novel human secret
C 23	369.5	7.5	190	19	AAW7092	Human RNA editing
24	369.5	7.5	190	21	AB111973	Human RNA editing
25	369.5	7.5	190	22	AA38905	Human polypeptide
26	369.5	7.5	210	22	AU23550	Novel human enzyme
27	369.5	7.5	221	22	AA40691	Human polypeptide
28	369.5	7.5	222	21	AA57061	Human prostate can
29	369.5	7.5	222	22	AU23537	Novel human enzyme
C 30	367.5	7.4	133	22	AU31783	Novel human secret
C 31	366.5	7.4	116	22	AU41736	Novel bone marrow
C 32	365	7.4	94	23	ABJ03693	Human ovary specif
C 33	364	7.4	209	22	AB12667	Human polypeptide
C 34	363	7.4	99	22	AB11839	Human novel protei
C 35	361	7.3	168	22	ABG7628	Novel human diagno
C 36	360.5	7.3	127	23	AB64678	Human albumin fusi
C 37	360.5	7.3	127	23	AB62080	Human gene 8 encod
C 38	359	7.3	144	23	AB79196	Human prostate spe
C 39	358.5	7.3	134	23	AA021703	Human secreted pro
C 40	358	7.3	100	23	ABP51849	Human colon specif
C 41	358	7.3	126	22	AA64466	Human gas vesicle
C 42	355	7.2	101	22	ABG01434	Novel human diagno
C 43	353.5	7.2	218	22	AAE11963	Human c-myb-relate
C 44	353	7.2	166	22	AU32028	Novel human secret
C 45	352	7.1	146	22	AB11580	Human secreted pro

ALIGNMENTS

RESULT 1  
AAB24198

ID AAB24198 standard; Protein; 198 AA.

XX

AC AAB24198;

XX 05-FEB-2001 (first entry)

DT

DE Human activation-induced cytidine deaminase SEQ ID NO:8.

XX

XX Activation-induced cytidine deaminase; AID: cytidine deaminase;

KW Immune related disease; allergy; allergic disease; anti-allergic;

KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;

KW gene therapy; B cell associated immune system disorder; food allergy;

KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;

KW IGA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;

KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;

KW ataxia telangiectasia; common variable immunodeficiency disorder;

KW major histocompatibility class II deficiency disease;

KW auto immunodeficiency syndrome; IgG subclass selection disorder.  
 OS Homo sapiens.  
 XX WO200058480-A1.  
 PN 05-OCT-2000.  
 XX 28-MAR-2000; 2000WO-JP01918.  
 XX 29-MAR-1999; 99JP-0087192.  
 PR 24-JUN-1999; 99JP-0178999.  
 PR 27-DEC-1999; 99JP-0371382.  
 XX (NLSB ) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 XX Honjo T, Muramatsu M;  
 PI WPI: 2000-611715/58.  
 XX N-PSDB; AAC55312.  
 DR Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 XX Claim 1: Page 140-141; 174pp; Japanese.  
 XX The present sequence is human activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
 CC antianemic, antisthmatic, ophthalmological, anti-HIV and  
 CC dermatological activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
 CC disease, IgA nephritis, gamma-globulinemia, atopic dermatitis, allergic  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
 CC immunodeficiency disorder, MHC (major histocompatibility class) class  
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
 CC encoding AID may be used for gene therapy and the antibodies to the AID  
 CC protein may be used for diagnosis and treatment of these disorders.  
 XX Sequence 198 AA;  
 XX  
 Alignment Scores:  
 Pred. No.: 5.3le-114 Length: 198  
 Score: 1086.00 Matches: 198  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 21.95% Indels: 0  
 DB: 21 Gaps: 0  
 US-09-966-880A-7 (1-2818) x AAB24198 (1-198)  
 QY 80 ATGACACGCTCTTGATGAACCGGAGGAGCTTTCTTACCAATTCAGAGGCGTGG 139  
 DB 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
 QY 140 GCTAAGGCTCGGCTGAGACCTACCTGTGTCTAGCTAGTGAAGAGCGGTGACAGTGTACA 199  
 DB 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
 QY 200 TCCTTTCTACGTGACCTTGTGTTATCTCGCAATTAAGACGGCTGCCAGTGGATGCTC 259  
 DB 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
 QY 260 TTCTCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACGCGTCACCTGG 319  
 DB 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80

QY 320 TTCACCTCTCTGGAGCCCTCTACGACTGTGTCCCGACATGTGCCGACTTTCTGCGAGGG 379  
 DB 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
 QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439  
 DB 101 AsnProAsnLeuSerLeuArgGlyPheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
 QY 440 GCTGAGCCGAGGCGTGGCGGCTGCACCGCGCGGTGCAAAATAGCATCATGACACC 499  
 DB 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
 QY 500 TTCAAAGATTATTTTACTCTGGAATACTTTTGTAGAAACCATGAAGAATTTCAAA 559  
 DB 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
 QY 560 GCCTGGGAGGCGTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGGCGCATCCTT 619  
 DB 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
 QY 620 TTGCGCCTGTATGAGTTGATGACTTACGAGACGCTTTCTGCTACTTTGGGACTT 673  
 DB 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198  
 RESULT 2  
 AAB24197  
 ID AAB24197 standard; Protein: 198 AA.  
 XX AAB24197;  
 AC AAB24197;  
 XX 05-FEB-2001 (first entry)  
 DT Mouse activation-induced cytidine deaminase SEQ ID NO:2.  
 XX  
 DE Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 XX Immune related disease; allergy; allergic disease; antiallergic;  
 KW antianemic; antisthmatic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disease;  
 KW major histocompatibility class II deficiency disease;  
 KW auto immunodeficiency syndrome; IgG subclass selection disorder.  
 XX Mus musculus.  
 OS  
 XX WO200058480-A1.  
 PN 05-OCT-2000.  
 XX 28-MAR-2000; 2000WO-JP01918.  
 XX 29-MAR-1999; 99JP-0087192.  
 PR 24-JUN-1999; 99JP-0178999.  
 PR 27-DEC-1999; 99JP-0371382.  
 XX (NLSB ) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 XX Honjo T, Muramatsu M;  
 PI WPI: 2000-611715/58.  
 XX N-PSDB; AAC55307.  
 DR Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 XX Claim 1: Page 131-132; 174pp; Japanese.  
 PS The present sequence is mouse activation-induced cytidine deaminase  
 XX (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC

CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
CC antianemic, antiasthmatic, ophthalmological, anti-HIV and  
CC dermatologic activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
CC disease, IgA nephritis, gamma-globulinemia, atopic dermatitis, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class)  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC IgG disorder, and IgG subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders.  
XX  
SQ Sequence 198 AA;

Alignment Scores:  
Pred. No.: 3.99e-105 Length: 198  
Score: 1008.00 Matches: 183  
Percent Similarity: 95.94% Conservative: 6  
Best Local Similarity: 92.89% Mismatches: 8  
Query Match: 20.37% Indels: 0  
DB: 21 Gaps: 0

US-09-966-880A-7 (1-2818) x AAB24197 (1-198)

QY 80 ATGACAGCTCTGTGATGACCGAGGAGTCTTCTTACCAATTCAAAATGCCGCTGG 139  
DB 1 MetAspSerLeuLeuMetLysGlnLysPheLeuTyrHisPheLysAsnValArgTrp 20  
QY 140 GCTAAGGGTGGCGCTGAGACCTACCTGTGCTAGTGAAGGCGTGACAGTGTACA 199  
DB 21 AlaLysGlyArgHisGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40  
QY 200 TCTTTTCTAGTACGCTTTGGTATCTTCGCAATAAGACGGCTGCCAGTGGAAATGCTC 259  
DB 41 SerCysSerLeuAspPheGlyHisLeuArgAsnLysSerGlyCysHisValGluLeu 60  
QY 260 TTCTCTCGCTACATCTCGACTGGACCTAGACCTCGCGCTGCTACCGCTCACTGG 319  
DB 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
QY 320 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGCTGCTGCGGCTTCTTCCGAGGG 379  
DB 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaGluPheLeuArgTrp 100  
QY 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTCTGTGAGGACCGAAG 439  
DB 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
QY 440 GCTGAGCCGAGGGCTGGCGGCTGCACCGCGCGGCTGCAAAATAGCATCATGACC 499  
DB 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleGlyIleMetThr 140  
QY 500 TTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAACCATGAAGACTTTCAAA 559  
DB 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnArgGluArgThrPheLys 160  
QY 560 GCCTGGAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGGCATCCTT 619  
DB 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuThrArgGlnLeuArgArgIleLeu 180  
QY 620 TTCCCTCTGATCAGCTGTGATGACTTACGAGAGCGCATTTCTGTTGGGA 670  
DB 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgMetLeuGly 197

RESULT 3

ID ABG11932

XX standard; Protein; 188 AA.

AC ABG11932;

XX

DT 18-FEB-2002 (first entry)  
XX Novel human diagnostic protein #11923.  
DE  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
XX N-PSDB; AAS76119.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 42291; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX

SQ Sequence 188 AA;

Alignment Scores:  
Pred. No.: 6.86e-38 Length: 188  
Score: 417.00 Matches: 99  
Percent Similarity: 64.40% Conservative: 24  
Best Local Similarity: 51.83% Mismatches: 41  
Query Match: 8.45% Indels: 28  
DB: 22 Gaps: 4

US-09-966-880A-7 (1-2818) x ABG11932 (1-188)

QY 2126 TTTTCTGAGACAGTCTTGTCTTGTCCCGAGCTGGAGTGCACGGCAC 2067  
DB 1 PheAsnPhePheLeuArgGlnSerLeuAlaLeuLeu-ProArgLeuGluCysSerGlyMe 20  
QY 2066 GATCTCAGCTTACTGCAACCTCCACCTCTGGGTTCAGCGATTCTCTCCCTCAGCCTC 2007  
DB 20 ttleSerAlaHisCysAsnLeuHisLeuProGlySerSerAsnSerProAlaLeuAlaPr 40

QY 206 CCAAGTAGCTGGGATTACAGGTGCTGCTACACGCCCTGGCTAAATTTTTCATTTTGGT 1947  
 Db :|||||  
 40 oargValAlaGlyIlelleSerThrCysHisAlaGlnValIlePheAlaCysLeuVa 60  
 1946 ACAGACGGGTTTTCCTGCTGGCCAGGCTGCTCAACCTCTGACACAGGTGATPC 1887  
 Db :|||||  
 60 lgiMetGlyPheHisAlaValGlyGlnAlaGlyLeuLysLeuThrSerGlyAspLe 80  
 1886 GCCCGCTGGCTGCCCTCCCA-AAAGTGTCTGGGATTACAGGCTGAGCCACACCGCCGCT 1828  
 Db :|||||  
 80 uproAlaLeuAlaSerGluLysCysTrpAspTyrArgGluGlnProArgProAlaLe 100  
 1827 CTCCTCTCTCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTT 1768  
 Db :|||||  
 100 u-----PhePheTrpTyrPheSerArgAspGlnArgPheThrThrLeuGlyGlnAl 117  
 1767 TGGCTTCAACACCTGCTTCAAGCCATCTCTCCCTCAACCTCCCAAGCAGCTAGAAC 1708  
 Db :|||||  
 117 aGlyLeuGluLeuLeuAlaSerSerAsnProProAlaSerAlaSerLysSerAlaGlyI 137  
 1707 TATGAGCTGTGCCAGGCTTCG-----CAGCTTTATCATCTGTT----- 1668  
 Db :|||||  
 137 eThrAspValSerHisGlyThrHisAlaLysArgLeuTyr-SerTrpLeuCysValSerG 157  
 1667 -----TTGTCAACATCTCTTATAGTTAA 1642  
 157 uThrAlaPheAsnProAsnPheSerThrPheArgLeuLysAsnIleLeuPheMetValS 177  
 1641 TATACCTTTGGGAGAGCATCATCACATAC 1611  
 Db :|||  
 177 nLeu-----GluHisAsnTyr 182

RESULT 4

ID ABB11478  
 AC ABB11478 standard; peptide; 126 AA.  
 XX ABB11478;  
 DT 11-JAN-2002 (first entry)  
 XX Human secreted protein homologue, SEQ ID NO:1848.  
 DE Human; cytokine; cell proliferation; tissue growth; immunomodulator; growth factor;  
 KW haematopoiesis regulation; tumour; cancer; tumour; haematopoietic disorder;  
 KW inhibin; chemotaxis; chemokinesis; thrombolytic; osteoporosis; arthritis;  
 KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;  
 KW myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;  
 KW chronic inflammatory condition; proliferative retinopathy;  
 KW atherosclerosis; coronary heart disease; arterial ischaemia;  
 KW bone disorder; osteoporosis; vascular growth disorder;  
 KW tissue regeneration; wound healing; infection; immune disorder;  
 KW cell culture; drug screening; gene therapy; anti-inflammatory;  
 KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;  
 KW cytosolic; osteopathic; vasotropic; cardiant; virucide; antibacterial;  
 KW antifungal; vulnary; antiulcer.

OS Homo sapiens.

XX W0200157188-A2.

PN 09-AUG-2001.

XX 05-FEB-2001; 2001WO-0503800.

XX 03-FEB-2000; 2000US-0496914.

PR 27-APR-2000; 2000US-0560875.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT;

PI WPI; 2001-457740/49.

DR

DR XX  
 PT Human proteins and DNA encoding sequences useful for preventing,  
 PT treating or ameliorating a medical condition in a mammalian subject  
 PT e.g. arthritis and cancer -  
 PS Claim 20; Page 197; 1963pp; English.  
 XX  
 CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and  
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The  
 CC invention also relates to vectors and recombinant host cells comprising a  
 CC nucleotide of the invention, methods of producing the novel polypeptides,  
 CC antibodies against the polypeptides, methods of detecting the nucleotides  
 CC or polypeptides in a sample, and methods of identifying compounds which  
 CC bind to polypeptides of the invention. Although novel, many of the  
 CC polypeptides of the invention have homology to known proteins, and hence  
 CC giving an insight into their probable biological activities, and hence  
 CC potential therapeutic applications. The polypeptides of the invention may  
 CC have various activities, including cytokine, cell proliferation or cell  
 CC differentiation activities; stem cell growth factor activity;  
 CC haematopoiesis regulatory activity; tissue growth activity;  
 CC immunomodulatory activity; activin- or inhibin-related activities;  
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or  
 CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC the invention on their biological activities, polypeptides and nucleotides of  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC vascular growth. Polypeptides involved with tissue regeneration and  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.  
 XX SQ Sequence 126 AA;

Alignment Scores:

Pred. NO.: 2.77e-36 Length: 126  
 Score: 402.00 Matches: 90  
 Percent Similarity: 72.39% Conservative: 7  
 Best Local Similarity: 67.16% Mismatches: 22  
 Query Match: 8.14% Indels: 15  
 DB: 22 Gaps: 3

US-09-966-880A-7 (1-2818) x ABB11478 (1-126)

QY 2183 AATGCTCCCATCTCTCTCT-----CCCAATATTTCTCTCTCTCTCTCTCT 2136  
 Db :|||||  
 2 AsnAlaSerThrValTyrSerSerGlnGlyAspProLys----- 14  
 QY 2135 CTCCTCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCTTTTTCCT 2076  
 Db :|||||  
 15 -----SerPhePhePheLeuLeuArgTrpSerLeuAlaLeuValAlaGlnAlaGlyGlu 32  
 QY 2075 CAACGGCAGCATCTCAGCTTACTGCAACCTCCACCTCCTGGTTCAGAGCATCTCTCGC 2016  
 Db :|||  
 33 Gln\*\*\*ArgAspLeuSerSerLeuGlnProProProGlyPheLys\*\*\*PheSerCys 52  
 QY 2015 CTCAGCCTCCCAAGTAGCTGGATTACAGGTGCTCTACACGCTGCTGCTACAGCTGCTAATTTTTC 1956  
 Db :|||||  
 53 LeuSerLeuProSerSerTrpAspTyrArgCysProLeuProCysLeuAlaAsn-Phe\*\* 72









CC clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1  
 CC are deposited with the American Type Culture Collection (ATCC) with  
 CC accession number 98599. The polynucleotides and encoded polypeptides have  
 CC cytotatic, anti-inflammatory, immunomodulator, vulnerary,  
 CC neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic  
 CC and anti-inflammatory activity and acting as cytokine modulators,  
 CC haematopoiesis regulators, tissue growth modulators and/or cadherin  
 CC suppressors. The polypeptides and polynucleotides are useful in gene  
 CC therapies, particularly for preventing, treating or ameliorating any of  
 CC the following diseases: immune deficiency and disorders: e.g. bacterial  
 CC or fungal infections, autoimmune disorders, cancer, systemic lupus  
 CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell  
 CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or  
 CC osteoarthritis; central and peripheral nervous system diseases and  
 CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's  
 CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;  
 CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis  
 CC or systemic inflammatory response syndrome, ischaemia-reperfusion  
 CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or  
 CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus  
 CC foliaceus.  
 XX  
 SQ Sequence 384 AA;

Alignment Scores:  
 Pred. No.: 1.23e-34 Length: 384  
 Score: 390.00 Matches: 83  
 Percent Similarity: 61.62% Conservative: 31  
 Best Local Similarity: 44.86% Mismatches: 59  
 Query Match: 7.88% Indels: 12  
 DB: 23 Gaps: 3

US-09-966-880A-7 (1-2818) x ABB55784 (1-384)

QY 95 ATGAACCGGAGGAGTTCTTTACCAATTCARAAATTCGCTGGCTAAGGTCGGCGT 154  
 Db 197 MetaspProThrPheThrPheAsnPheAsnGluProTrpValArgGlyArgHis 216  
 QY 155 GAGACCTACCTGTGCTACGTAGTGAAGAGCGCTGACAGTCTACATCTTTTACATGGAC 214  
 Db 217 GluThrTyLeuCysTyrGluValGluArgMetHisAsnAspThrTrpValLeuLeuAsn 236  
 QY 215 -----TTTGGTTATCTTCGAATAAG-----AACGGCTGC 244  
 Db 237 GlnArgArgGlyPheLeuCysAsnGlnAlaProHisLysHisGlyPheLeuGluGlyArg 256  
 QY 245 CAGTCGGAATTCCTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGC 304  
 Db 257 HisAlaGluLeuCysPheLeuAspValIleProPheTrpLysLeuAspGlnAsp 276  
 QY 305 TACCGCGTCACCTGGTTACCTCCTCGAGCCCTCTGCTACGACTGTGCCGACATGTGGC 364  
 Db 277 TyrArgValThrCysPheThrSerTrpSerProCysPheSerCysAlaGlnGluMetAla 296  
 QY 365 GACTTCTCGGAGGAGACCCACCTCAGTCTGAGGATCTTACCGCGGCTCTACTTC 424  
 Db 297 LysPheIleSerLysAsnLysHisValSerLeuCysIlePheThrAlaArgIleTrp--- 315  
 QY 425 TGTGAGGACCGAAGGCTGAGCGGCTGCGGCTGCACCGCGCGGGTCA 484  
 Db 316 ---AspAspGlnGlyArgCysGlnGluGlyLeuArgThrLeuAlaGluAlaLys 334  
 QY 485 ATAGCATCATGACCTTCAAGATATTTTACTGCTGGAATCTTTTGTAGAAACCAT 544  
 Db 335 IleSerIleMetThrTrpSerGluPheLysHisCysTrpAspThrPheValAspHisGln 354  
 QY 545 GAAGAACCTTCAAGCCCTGGAAGGCTGCATGAAATTCAGTCTGCTCTCCAGAC 604  
 Db 355 GlyCysProPheGlnProTrpAspGlyLeuAspGluHisSerGlnAspLeuSerGlyArg 374  
 QY 605 CTTGGCGCATCTT 619  
 Db 375 LeuArgAlaIleLeu 379

RESULT 9  
 AAM83429  
 ID AAM83429 standard; Protein; 119 AA.  
 XX  
 AC AAM83429;  
 XX  
 DT 07-NOV-2001 (first entry)  
 XX  
 DE Human immune/haematopoietic antigen SEQ ID NO:11022.  
 XX  
 KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
 KW cytostatic; gene therapy; vaccine; metastasis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200157182-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01354.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180828.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226868.  
 PR 23-AUG-2000; 2000US-0227182.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.

PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239335.  
 PR 13-OCT-2000; 2000US-0239337.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.

PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251858.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 DR WPI; 2001-483426/52.  
 DR N-PSDB; AAK56210.  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 PT useful for preventing, diagnosing and/or treating cancers and  
 PT metastasis -  
 XX  
 PS Claim 11; SEQ ID NO 11022; 3071pp + Sequence Listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)  
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patients own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting  
 CC the nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK87694 represent human immune/haematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
 CC represent sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 119 AA;  
 Alignment Scores:  
 Pred. No.: 1.05e-34 Length: 119  
 Score: 388.00 Matches: 78  
 Percent Similarity: 82.00% Conservative: 4  
 Best Local Similarity: 78.00% Mismatches: 17  
 Query Match: 7.84% Indels: 1  
 DB: 22 Gaps: 0  
 US-09-966-880A-7 (1-2818) x AAM83429 (1-119)  
 QY 1832 CCGGGGTGGTGGTCCAGCGCTGTAATCCAGCACCTTTGGAGGCGCGGCGGATC 1891  
 Db 21 ProGlySerVal-LeuThrProValIleProThrLeuThrProGluAlaGlyGlySe 40  
 QY 1892 ACCTGTGGTCCAGGAGTTTGAGACAGCCTGGCCAAACATGGCAACCCCGTCTGTACTCA 1951  
 Db 40 rProGluValArgSerArgProAlaTrpProThrTrpGlyAsnLeuPheSerThrLy 60  
 QY 1952 AAATGCAAAATTTAGCCAGCGGTGGTAGCAGGACCTGTATCCAGCTACTTGGGAGGC 2011  
 Db 60 sAsnThrLysIleSerArgAlaTrpTrpGlnLeuProValIleProAlaThrGlnGluAl 80  
 QY 2012 TGAGGCGAGGAATCGCTTGAACCCAGGAGGTGGAGTTCAGTAAGTCGATCGTCC 2071  
 Db 80 aGluAlaGlyGluSerLeuGluProArgTrpArgLeuGlnTrpAlaLysValAlaPr 100  
 QY 2072 GTTGCACTCCAGCGTGGGCGACAGCAAGACTCTCTCAGAAAAAAGAAAAA 2129

Db 100 oLeuHisSerLeuGlySerLysSerLysThrSerSerGlnLysLysLysLys 119  
 AAY84437

RESULT 10  
 ID AAY84437 standard; Protein; 384 AA.  
 AC AAY84437;

XX 25-JUL-2000 (first entry)  
 XX Amino acid sequence of a human RNA-associated protein.

XX Human; RNA-associated protein; cell proliferation; cancer; inflammation;  
 XX immune response; reproductive disorder; actinic keratosis;  
 KW atherosclerosis; arteriosclerosis; bursitis; cirrhosis; hepatitis;  
 KW mixed connective tissue disease; myelofibrosis; primary thrombocytopenia;  
 KW paroxysmal nocturnal hemoglobinuria; polycythemia vera; psoriasis;  
 XX trauma.

XX Homo sapiens.

Key	Location/Qualifiers
Modified-site 13	/note= "potential phosphorylation site"
Modified-site 19	/note= "potential phosphorylation site"
Modified-site 28	/note= "potential phosphorylation site"
Modified-site 32	/note= "potential phosphorylation site"
Modified-site 86	/note= "potential phosphorylation site"
Modified-site 148	/note= "potential phosphorylation site"
Modified-site 167	/note= "potential phosphorylation site"
Modified-site 277	/note= "potential phosphorylation site"
Modified-site 311	/note= "potential phosphorylation site"
Modified-site 327	/note= "potential phosphorylation site"
Modified-site 349	/note= "potential phosphorylation site"
Modified-site 372	/note= "potential phosphorylation site"
Modified-site 427	/note= "potential phosphorylation site"

WO200015799-A2.

23-MAR-2000.

17-SEP-1999; 99WO-US21688.

17-SEP-1998; 98US-0156039.

22-SEP-1998; 98US-0158720.

04-NOV-1998; 98US-0186815.

08-APR-1999; 99US-0128660.

(INCY-) INCYTE PHARM INC.

Tang YT, Corley NC, Guegler KJ, Gorgone GA, Patterson C;  
 Hillman JL, Baughn MR, Lal P, Azimzai Y, Yue H, Yang J;

WPI; 2000-271437/23.

N-PSDB; AAA12409.

XX New polypeptides and polynucleotides, useful for preventing and  
 PT treating a disorder associated with increased or decreased expression  
 PT of RNA associated proteins -

XX

PS Claim 1; Page 101-102; 131pp; English.  
 XX The present sequence represents a human RNA-associated protein. The  
 CC expression of RNA-associated proteins is closely associated with  
 CC reproductive tissues, nervous tissues, cell proliferation including  
 CC cancer, inflammation and immune responses, and so they may be used  
 CC for diagnosis, treatment or prevention of cell proliferative,  
 CC immune/inflammatory disorders, and reproductive disorders. Diseases  
 CC and disorders which may be treated include actinic keratosis,  
 CC atherosclerosis, arteriosclerosis, bursitis, cirrhosis, hepatitis,  
 CC mixed connective tissue disease, myelofibrosis, primary thrombocytopenia  
 CC hemoglobinuria, polycythemia vera, psoriasis, paroxysmal nocturnal  
 CC and cancers, and trauma.

SQ Sequence 384 AA;

Alignment Scores:  
 Pred. No.: 2,07e-34 Length: 384  
 Score: 388.00 Matches: 82  
 Percent Similarity: 61.62% Conservative: 32  
 Best Local Similarity: 44.32% Mismatches: 59  
 Query Match: 7.84% Indels: 12  
 DB: 21 Gaps: 3

US-09-966-880A-7 (1-2818) x AAY84437 (1-384)

QY	95	ATGAACGGGAGGAGTTCTTTTACCAATTTACAAATGTCGGCTGAGGTCGGCGT	154
Db	197	MetAspProProThrPheThrPheAsnGluProTrpValArgGlyArgHis	216
QY	155	GAGACCTACCTGTGCTAGTGTAGAGAGCGCTGACAGTCTACATCTTTTACCTGGAC	214
Db	217	GluThrTyrLeuCysTyrGluValGluArgMetHisAsnAspThrTrpValLeuLeuAsn	236
QY	215	-----TTTGGTTATCTTCGAATAAG-----AACGGCTGC	244
Db	237	GlnArgArgGlyPheLeuLeuGlnAlaProHisLysHisGlyPheLeuGluArg	256
QY	245	CACGTGGAATGTCTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGC	304
Db	257	HisAlaGluLeuLeuGluLeuAspValIleProPheTrpLysLeuAspGlnAsp	276
QY	305	TACCGCTCACCTGTGTTTACCTCTCGAGCCCTCTACGACTGTGCCGACATGTGCC	364
Db	277	TyrArgValThrCysPheThrSerTrpSerProCysPheSerCysAlaGlnMetala	296
QY	365	GACTTCTGCGAGGAGCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTC	424
Db	297	LysPheIleSerLysAsnLysHisValSerLeuCysIlePheThrAlaArgIleTrp	315
QY	425	TGTGAGGACCGCAAGCTGAGCCCGGCTGCGGCTGCGCCGCGGGGTCAA	484
Db	316	---AspAspGlnGlyArgCysGlnGluGlyLeuArgThrLeuAlaGluAlaGlyAlaLys	334
QY	485	ATAGCCATCATGACCTTCAAAGATTATTTTACTGTGGAATACTTTTGTAGAAACCAT	544
Db	335	IleSerIleLeuThrTyrSerGluPheLysHisCysTrpAspThrPheValAspHisGln	354
QY	545	GAAGAACTTTCAAAGCCTGGAAGGCTGCATCAAAATTCAGTCTCTCTCCAGACAG	604
Db	355	GlyCysProPheGlnProTrpAspGlyLeuGluHisSerGlnAlaLeuSerGlyArg	374
QY	605	CTTGGCGCATCTCTT	619
Db	375	LeuArgGlyIleLeu	379

RESULT 11  
 AAU31902  
 ID AAU31902 standard; Protein; 107 AA.  
 XX AAU31902;  
 AC AAU31902;  
 XX 18-DEC-2001 (first entry)  
 DT

Db	63	uValGluThrGlyPheHisValGlyGlnAlaGlyLeuGluLeuLeuThrSerGlyAs	83
Qy	1889	TCGCGCGGCTCGCGCTCCCAAAGTGTGGATTACAGGCGTGAGCCACACGCGCCGCC	1830
Db	83	pProAlaSerAlaSerGlnSerValGlyIleThrGlyValSerHisArgAlaArgPr	103
Qy	1829	CTCTCTC 1823	
Db	103	oGluLeu 105	
RESULT 12			
ABBL2093	ABBL2093 standard; peptide; 134 AA.		
ID			
AC			
XX	ABBL2093;		
XX			
DT	11-JAN-2002 (first entry)		
XX			
DE	Human secreted protein homologue, SEQ ID NO:2463.		
XX			
KW	Human; cytokine; cell proliferation; cell differentiation; growth factor; hematopoiesis regulation; tissue growth; immunomodulator; activin; inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis; proliferation; metastasis; cancer; tumour; haematopoietic disorder; myeloid cell disorder; lymphoid cell disorder; asthma; arthritis; chronic inflammatory condition; proliferative retinopathy; atherosclerosis; coronary heart disease; arterial ischaemia; bone disorder; osteoporosis; vascular growth disorder; tissue regeneration; wound healing; infection; immune disorder; cell culture; drug screening; gene therapy; antiinflammatory; antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic; cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial; antifungal; vulnery; antiulcer.		
XX			
OS	Homo sapiens.		
XX			
PN	W0200157188-A2.		
XX			
PD	09-AUG-2001.		
XX			
PF	05-FEB-2001; 2001WO-US03800.		
XX			
PR	03-FEB-2000; 2000US-0496914.		
PR	27-APR-2000; 2000US-0560875.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Drmanac RT;		
XX			
DR	WPI; 2001-457740/49.		
DR	N-PSDB; ABA09337.		
XX			
PT	Human proteins and DNA encoding sequences useful for preventing, treating or ameliorating a medical condition in a mammalian subject		
PT	e.g. arthritis and cancer		
XX			
PS	Claim 20; Page 306; 1963pp; English.		
XX			
CC	Sequences ABBL0981-ABBL2330 represent 1350 novel human polypeptides, and sequences ABA08225-ABA09574 represent nucleic acids encoding them. The invention also relates to vectors and recombinant host cells comprising a nucleotide of the invention, methods of producing the novel polypeptides, antibodies against the polypeptides, methods of detecting the nucleotides or polypeptides in a sample, and methods of identifying compounds which bind to polypeptides of the invention. Although novel, many of the polypeptides of the invention have homology to known proteins, thereby giving an insight into their probable biological activities, and hence potential therapeutic applications. The polypeptides of the invention may have various activities, including cytokine, cell proliferation or cell differentiation activities; stem cell growth factor activity; hematopoiesis regulatory activity; tissue growth activity; immunomodulatory activity; activin- or inhibin-related activities; chemotactic or chemokinetic activities; haemostatic, thrombotic or		

CC thrombolytic activities; receptor or ligand activities; or may be  
 CC involved in oncogenesis, cancer cell proliferation or metastasis.  
 CC Depending on their biological activities, polypeptides and nucleotides of  
 CC the invention are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g., by protein or gene therapy. Such conditions include  
 CC cancers, hematopoietic disorders (e.g., myeloid or lymphoid cell  
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),  
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,  
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal  
 CC repair (or nucleic acids encoding them) may be used to promote wound  
 CC healing (e.g., of burns, incisions and ulcers), while those with  
 CC immunomodulatory activities may be used in the treatment of viral,  
 CC bacterial and fungal infections in addition to immune disorders.  
 CC Polypeptides with growth factor activity may be used in cell cultures to  
 CC promote cell growth. For example, such polypeptides may be used to  
 CC manipulate stem cells in culture to give rise to neuroepithelial cells  
 CC that can be used to augment or replace cells damaged by illness,  
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides  
 CC may also be used in the diagnosis of the above conditions, and in drug  
 CC screening techniques. The present sequence represents a novel human  
 CC polypeptide of the invention.  
 XX  
 SQ Sequence 134 AA;

Alignment Scores:  
 Pred. No.: 3,65e-34 Length: 134  
 Score: 383.50 Matches: 88  
 Percent Similarity: 71.64% Conservative: 8  
 Best Local Similarity: 65.67% Mismatches: 31  
 Query Match: 7.77% Indels: 7  
 DB: 22 Gaps: 2

US-09-966-880A-7 (1-2818) x ABB12093 (1-134)

QY 2117 TTCTGAGACAGATCTTCTCTTCTGCGCCAGGTGGAGTCAACGGCAGCATCTCAGC 2058  
 Db 1 PheLeuArgProSerPheAlaLeuValProGlnAlaGlyValGlnTrpCysAlaLeuSer 20  
 QY 2057 TTACTGCACTCCACCTCCTCGGTTCACAGCGATTCTCTGCTCAGCTCCCAAGTAGC 1998  
 Db 21 TrpLeuGlnProSerProArgPheLys\*\*\*PheSerCysLeuSerLeuProSerSer 40  
 QY 1997 TGGATTACAGGTGCTGTACACGCTGCTGCTAAT---TTTGCATTTTGGTACAGAC 1941  
 Db 41 TrpAspTyrArgHisValProArgProAlaAsnPhePheValLeuLeu-ValGluTh 60  
 QY 1940 GGGGTTTCCCATGTGGCCAGGTGGTCTCAAACTCTGTACACAGGTGATCGCCCGG 1881  
 Db 60 rGlyPheLeuHisValGlyGlnAlaGlyHisGluProLeuThrSerGlyAspProAl 80  
 QY 1880 CTGCGCTCCCAAGTGTGGGATTACAGCGGTGAGCCACGCGCCCTCTCTCTT 1821  
 Db 80 aSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisGlnAlaTrpProSer----- 98  
 QY 1820 TCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 1761  
 Db 99 -----Phe-PheIlePheSerArgAspThrValLeuLeuCysSerGlyTrpSera 116  
 QY 1760 GAACACCTGTGTTCAGACCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1723  
 Db 116 rgThrSerGlyLeuGlyGlnSerAlaCysLeuSerLeu 128

RESULT 13  
 AAB90650  
 ID AAB90650 standard; Protein; 239 AA.  
 XX  
 AC AAB90650;

01-JUN-2001 (first entry)

Human secreted protein, SEQ ID NO: 193.

KW Human; secreted protein; immunomodulatory; antisclerotic;  
 KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;  
 KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;  
 KW neotropic; anticonvulsant; antialzheimers; antiparkinsonian;  
 KW antimicrobial; vulnery; vaccine; gene therapy; cancer;  
 KW protein coordinate data; infection.  
 XX Homo sapiens.  
 OS  
 XX WO200121658-A1  
 PN  
 XX 29-MAR-2001.  
 PD  
 XX 22-SEP-2000; 2000WO-US26013.  
 PF  
 XX 24-SEP-1999; 99US-0155709.  
 PR  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;  
 PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;  
 PI Young PE, Wei P, Florence KA;  
 XX  
 XX WPI; 2001-235311/24.  
 DR  
 XX Nucleic acids encoding 32 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
 PT disease and diabetic retinopathy -  
 PT  
 XX Disclosure; Page 858; 890pp; English.

CC The present sequence is provided in a specification relating to nucleic  
 CC acid molecules encoding 32 novel human secreted polypeptides. The nucleic  
 CC acid molecules and polypeptides may be used in the prevention, diagnosis  
 CC and treatment of diseases such as immune disorders (e.g. multiple  
 CC sclerosis, systemic lupus erythematosus and human immuno-deficiency virus  
 CC (HIV) infections), hyperproliferative disorders (e.g. cancers and  
 CC Gaucher's disease), cardiovascular diseases (e.g. cancers and  
 CC Chaga's cardiomyopathy and coronary arteriosclerosis), angiotonic  
 CC disorders (e.g. corneal graft neovascularisation and diabetic  
 CC retinopathy), neurological disorders (e.g. Huntington's chorea,  
 CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or  
 CC for promoting wound healing, regeneration and/or chemotaxis. The nucleic  
 CC acid molecules may be used to produce the secreted polypeptides. The nucleic  
 CC also be used as DNA probes in diagnostic assays to detect and quantitate  
 CC the presence of similar nucleic acid sequences in samples. The  
 CC polypeptides may be used as antigens in the production of antibodies and  
 CC in assays to identify modulators of their expression and activity.

SQ Sequence 239 AA;

Alignment Scores:  
 Pred. No.: 8,63e-34 Length: 239  
 Score: 381.50 Matches: 87  
 Percent Similarity: 72.39% Conservative: 10  
 Best Local Similarity: 64.93% Mismatches: 27  
 Query Match: 7.73% Indels: 10  
 DB: 22 Gaps: 3

US-09-966-880A-7 (1-2818) x AAB90650 (1-239)

QY 2189 CCTTGAATGCTTCCCA-----TCTTCTCTCTCCCAATATTGTCCTT 2145  
 Db 87 ProCysCysSerSerSerAlaTrpProGluGlySerPheArgProPheGlnMeCASHLeu 106  
 QY 2144 TCTCTCTCTCTCTCTCTT-----TTTTTTTTTTTCTGAGACAGCTCTTTCCTT 2094  
 Db 107 PheSerPheLeuSerPhePhePheLeuPhePheLeuArgTrpSerLeuThr-Le 126  
 QY 2093 GTCGCCAGGTGGAGTCAACGGCAGCATCTACGTACTGCAACCTCCACCTCTCTGGG 2034  
 Db 126 UserProArgLeuGluCysSerSerAlaIleSerAlaHisCysAsnLeuArgLeuProG 146



QY 2033 TTCAACGATTCTCTGCTCAGCCCTCCCAAGTAGCTGGATTACAGGTGCTCTGCTACCA 1974  
 Db 146 ySerSerAsnSerProAlaLeuAlaSerGlnValAlaGlyIleThrGlyIleCysHisH 166  
 QY 1973 CGCCTGGCTAATTTTGGCATTTTGAGTACAGACGGGTTTGGCATGTTGGCAGGCTGG 1914  
 Db 166 sAlaArgGlnIlePheValPheLeuValGluThrGlyPheCysHisValGlyGlnAlaG 186  
 QY 1913 TCTCAACCTCCTGACACAGGTGATCCCGCGCTCGGCTCCCAAGTCTGGGATTAC 1854  
 Db 186 yLeuGluLeuLeuIleSerGlyAspSerProAlaSerAlaPheGlnSerAlaGlyIleI 206  
 QY 1853 AGCGTGAGCCACCGCCGCGCC---TCTCTCTTCTT 1817  
 Db 206 eGlyValSerHisArgAlaArgProGlySerValPheLeu 219  
 RESULT 14  
 AAB90647  
 ID AAB90647 standard; Protein; 310 AA.  
 AC AAB90647;  
 XX  
 DT 01-JUN-2001 (first entry)  
 XX  
 DE Human secreted protein, SEQ ID NO: 190.  
 KW Human; secreted protein; immunomodulatory; antisclerotic;  
 KW dermatological; antiinflammatory; anti-HIV; cytostatic; cardiant;  
 KW vascular; anti-angiogenic; ophthalmological; neuroprotectant;  
 KW neotropic; anticonvulsant; antitumor; antiparkinsonian;  
 KW antimicrobial; vulnary; vaccine; gene therapy; cancer;  
 KW protein coordinate data; infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200121658-A1.  
 XX  
 PD 29-MAR-2001.  
 XX  
 PF 22-SEP-2000; 2000WO-US26013.  
 XX  
 PR 24-SEP-1999; 99US-0155709.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Ni J, Baker KP, Birse CE, Ebner R, Fiscella M, Komatsoulis GA;  
 PI Lafleur DW, Moore PA, Olsen HS, Rosen CA, Ruben SA, Soppet DR;  
 PI Young PE, Wei P, Florence KA;  
 XX  
 WI 2001-235311/24.  
 XX  
 DR Nucleic acids encoding 32 human secreted polypeptides, useful for  
 PT preventing, diagnosing and/or treating e.g. cancers, Parkinson's  
 PT disease and diabetic retinopathy -  
 XX  
 PS Disclosure; Page 856; 890pp; English.  
 XX  
 CC The present sequence is provided in a specification relating to nucleic  
 CC acid molecules encoding 32 novel human secreted polypeptides. The nucleic  
 CC acid molecules and polypeptides may be used in the prevention, diagnosis  
 CC and treatment of diseases such as immune disorders (e.g. multiple  
 CC sclerosis, systemic lupus erythematosus and human immuno-deficiency virus  
 CC (HIV) infections), hyperproliferative disorders (e.g. cancers and  
 CC Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome,  
 CC Chaga's cardiomyopathy and coronary arteriosclerosis), angioleptic  
 CC disorders (e.g. corneal graft neovascularisation and diabetic  
 CC retinopathy), neurological disorders (e.g. Huntington's chorea,  
 CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or  
 CC for promoting wound healing, regeneration and/or chemotaxis. The nucleic  
 CC acid molecules may be used to produce the secreted polypeptides. They may  
 CC also be used as DNA probes in diagnostic assays to detect and quantitate  
 CC the presence of similar nucleic acid sequences in samples. The  
 CC polypeptides may be used as antigens in the production of antibodies and

CC in assays to identify modulators of their expression and activity.  
 XX  
 SQ Sequence 310 AA;  
 Alignment Scores:  
 Pred. No.: 1e-33 Length: 310  
 Score: 381.50 Matches: 87  
 Percent Similarity: 72.39% Conservative: 10  
 Best Local Similarity: 64.93% Mismatches: 27  
 Query Match: 7.73% Indels: 10  
 DB: Gaps: 3  
 US-09-966-880A-7 (1-2818) x AAB90647 (1-310)  
 QY 2189 CTTTGAATGCTTCCCA-----TCCTTCTCTCCAAATATTGTTCTCTT 2145  
 Db 158 ProCysCysSerSerSerAlaTrpProGluGlySerPheArgPropheGlnMetAsnLeu 177  
 QY 2144 TCTCTCTCTCTCTCTCTT-----TTTTTTTTTCTGAGACAGAGTCTTCTCTT 2094  
 Db 178 PheSerPheLeuSerPhePhePhePhePhePhePhePhePhePhePhePhePhePhePhe 197  
 QY 2093 GTCGCCAGCTGGAGTGCAACGCGAGGATCTCAGCTTACTGCAACCTCCACCTCTCTGGG 2034  
 Db 197 uSerProArgLeuGluCysSerSerAlaIleSerAlaHisCysAsnLeuArgLeuProG 217  
 QY 2033 TTCAACGATTCTCTGCTCAGCTCCCAAGTAGCTGGATTACAGGTGCTCTGCTACCA 1974  
 Db 217 ySerSerAsnSerProAlaLeuAlaSerGlnValAlaGlyIleThrGlyIleCysHisH 237  
 QY 1973 CGCCTGGCTAATTTTGCATTTTGTAGTACAGCGGGGTTTGGCATGTTGGCCAGGCTGG 1914  
 Db 237 sAlaArgGlnIlePheValPheLeuValGluThrGlyPheCysHisValGlyGlnAlaG 257  
 QY 1913 TCTCAACCTCTCAGCACAGGTGATCGCGCGCTCGGCTCCCAAGTCTCCAGGATTAC 1854  
 Db 257 yLeuGluLeuLeuIleSerGlyAspSerProAlaSerAlaPheGlnSerAlaGlyIleI 277  
 QY 1853 AGCGTGAGCCACCGCCGCGCC---TCTCTCTTCTT 1817  
 Db 277 eGlyValSerHisArgAlaArgProGlySerValPheLeu 290  
 RESULT 15  
 AAB42917  
 ID AAB42917 standard; Protein; 165 AA.  
 AC AAB42917;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF3681 polypeptide sequence SEQ ID NO:5362.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnary; antipsoriatic; antiparkinsonian; neotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;  
 KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;  
 KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;  
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200058473-A2.  
 XX  
 PD 05-OCT-2000.  
 XX



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 14, 2003, 18:12:33 ; Search time 81.1094 Seconds

(without alignments)  
6680.033 Million cell updates/sec

Title: US-09-966-880A-7

Perfect score: 4948

Sequence: 1 agagaccattacattatga.....aaaaaaaaaaaaaaaaaa 2818

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Fgapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 9613422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q/cgn21/USPTO\_spool/US09966880/runat\_14062003\_175525\_10329/app\_query.fasta\_1.9493  
-DB=PIR\_73 -QMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USPR=US09966880.ecgn\_1.1.456.0runat\_14062003\_175525\_10329 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	370	7.5	673	4 F40201	artifact-warning s
c 2	348	7.1	613	4 C40201	artifact-warning s
c 3	340	6.9	627	4 A40201	artifact-warning s
4	339	6.9	613	4 C40201	artifact-warning s
5	330	6.7	673	4 F40201	artifact-warning s
6	329.5	6.7	627	4 A40201	artifact-warning s
c 7	280.5	5.7	301	4 B40201	artifact-warning s
c 8	270	5.5	597	4 E40201	artifact-warning s
9	265.5	5.4	597	4 E40201	artifact-warning s
10	254	5.1	196	2 I38022	hypothetical prote
11	246.5	5.0	301	4 B40201	phorbollin i - huma
12	245	5.0	116	2 G01233	x-linked retinopat
c 13	243	4.9	100	2 A46010	apolipoprotein B m
14	225	4.5	236	2 A53853	

15	223.5	4.5	418	2	S41044	chromosomal protei
16	221	4.5	236	2	J59323	apolipoprotein B m
17	213.5	4.3	229	2	JC4269	apolipoprotein B m
18	213.5	4.3	229	2	I48249	apolipoprotein B m
19	199.5	4.0	229	2	I59577	apolipoprotein B m
c 20	186	3.8	579	4	D40201	artifact-warning s
c 21	185.5	3.8	499	2	S65657	alpha-lc-adrenergi
c 22	182.5	3.7	841	1	I78885	serine/threonine-s
c 23	171	3.5	579	4	D40201	artifact-warning s
c 24	153	3.1	46	2	I54375	gene NF2 protein -
c 25	149.5	3.0	114	2	JC5238	galactosylceramide
c 26	148	3.0	39	2	I54374	gene NF2 protein -
c 27	143.5	2.9	124	2	A47582	B-cell growth fact
c 28	140	2.8	79	2	A56194	thromboxane A-2 re
c 29	140	2.8	407	2	T02670	probable thromboxa
c 30	134	2.7	447	2	A57034	transcription fact
c 31	130	2.6	252	2	PC4259	ferritin associate
c 32	130	2.6	369	2	A53959	thromboxane A-2 re
c 33	126.5	2.6	342	2	PC4211	hepatocellular car
c 34	120	2.4	53	2	A42442	integrin beta-1 ch
c 35	120	2.4	619	2	A60646	transforming prote
c 36	108	2.2	1191	2	S35305	zinc finger protei
c 37	106	2.1	773	2	D90099	hypothetical prote
c 38	103.5	2.1	536	2	E72073	hypothetical prote
c 39	103.5	2.1	536	2	E86550	hypothetical prote
c 40	102.5	2.1	655	2	S59836	hypothetical prote
c 41	101.5	2.1	295	2	S58850	homeotic protein a
c 42	98.5	2.0	277	2	A46241	interferon respons
c 43	98.5	2.0	369	2	S63464	hypothetical prote
c 44	98.5	2.0	731	2	A99106	hypothetical prote
c 45	96.5	2.0	228	2	AD2935	hypothetical prote

ALIGNMENTS

RESULT 1

F40201  
artifact-warning sequence (translated ALU class F) - human  
C;Species: Homo sapiens (man)  
C;Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000  
C;Accession: F40201  
R;Claverie, J.M.  
personal communication, 1992  
A;Reference number: A40201  
A;Accession: F40201  
A;Molecule type: DNA  
A;Residues: 1-673 <CLA>  
R;Claverie, J.M.  
Genomics 12, 838-841, 1992  
A;Title: Identifying coding exons by similarity search: Alu-derived and other potent  
A;Reference number: A40200; MUID:92241891; PMID:1572661  
A;Contents: annotation  
C;Comment: This "warning" entry is a conceptual translation in all 6 reading frames  
in-frame stop codons are shown as 'X'.  
C;Comment: Any significant similarity of a predicted protein sequence to a portion o

Alignment Scores:			
Pred. No.:	2.41e-27	Length:	673
Score:	370.00	Matches:	115
Percent Similarity:	49.81%	Conservative:	16
Best Local Similarity:	43.73%	Mismatches:	75
Query Match:	7.50%	Indels:	59
DB:	4	Gaps:	8

US-09-966-880A-7 (1-2818) x F40201 (1-673)

QY	2299	GAATATTTTCTCACTGTCAAAGACACTGTTAGGACAAT-----AGACACAA	2249
		::: :::::	
Db	370	GlnValLeuAlaHisCysSerLeuAsnLeuGlySerSerProAlaSerVal	389
QY	2248	AAGAGACAAATAGGATCCCT-----TATTGGCTCCT-TACATTTTGTG	2205



of sequence, and significant similarity of a predicted protein sequence to a protein

1. The first step is to identify the problem or question that needs to be answered. This involves understanding the context and the specific requirements of the task.

C;Comment: Any significant sl

Alignment Scores:

Pred. No.: 2e-23 Length: 673  
Score: 330.00 Matches: 98  
Percent Similarity: 52.85% Conservative: 32  
Best Local Similarity: 39.84% Mismatches: 59  
Query Match: 6.67% Indels: 58  
DB: 4 Gaps: 6

US-09-966-880A-7 (1-2818) x F40201 (1-673)

```
QY 1715 CTGCTGGAGGCTTGGAGGAGGATGCTTGAACACACAGGTGTTCAAGGC----- 1765
Db 57 LeuHisGlyArgGlnArgGlnGluAsnArgLeuAsnProGlyGly***GlyCysSerGlu 76
QY 1766 -----CAGCCTGGGCAACATAACAAGATCCTGCTCTCAAAAAA 1804
Db 77 ProLysLeuAlaThrAlaLeuProGlyCysGlnSerLysGlyLeuSerGlnLysGln 96
QY 1805 AAAAAAAAAAAGAACAGAGAGG-----GCCGGGCGT 1839
Db 97 LysGlnSerLysLysLysLysLysLysThrProLysAsnLys*****AlaGlyCys 116
QY 1840 GTGGCTCAGCGCTGTAATCCAGCACTTTGGAGCGCGAGCGGCGGATCACCTGTGG 1899
Db 117 GlyGlySerArgLeuSerSerGlnHisPheGlyArgPro-GlyGlyGlnIleThr***cl 136
QY 1900 TCAGGAGTTTGACACAGCCTGGCCAAATGCAAAACCCCTCTGTACTCAAAATGCAA 1959
Db 136 yGlnGluPheGluThrSerLeuIleAsnMetValLysLeuCysLeuTyr**LysTyrI 156
QY 1960 AAATTAGCAGCGGTGTAGCAGGCACCTGTATCCACGCTACTTGGGAGGCTGAGGCA 2019
Db 156 easn***ProGlyMetAlaAlaHisAlaCysAsnProSerTyrThrGlyAspArgGlyAr 176
QY 2020 GAGATTCGCTTGAACCCAGGAGGTGGAGTTCAGTAAGCTGAGATCGTGCCTTGCAC 2079
Db 176 gargileAla**ThrGlnGluValGluAlaValSer-GlnAsnLeuProLeuHisS 196
QY 2080 CAGCCTGGGACAGACAGACAGCTCTCTCAGAAAAA*****AAGAGAGAGAG 2139
Db 196 exSerArgGly-ValArgAlaArgAlaTyrLeuLysAsnLysAlaLysLysLys 215
QY 2140 AGAGAAGA----- 2148
Db 216 LysLysLysProProLysThrLys*****LeuGlyValValAlaHisAlaCysHis 235
QY 2149 GAACAATATTGGAGAGAGGATGGGAAGCATTTGCAAGGAAATGTGCTTTATCCAA 2208
Db 236 ProSerThrLeuGlyAspGlnGluGlyArgSerLeuGluValArg----- 250
QY 2209 AAAATGAAGGAGCCCAATAAGGATCCCTATTGTCTCTTTTGGTGCTATTGTGCC 2268
Db 251 -----SerLeuArgProAla**SerThrTrp***AsnCys 262
QY 2269 ACAACTGCTTTCACAGTCAGAAAAAT-----ATTCAGAATAACCATATCCCTGT 2322
Db 263 ValSerIle-----LysAsnThr***IleSerGlnGluTrpArgProMetPro 278
QY 2323 TTATTACCTAGC 2334
Db 279 ValIleProAla 282
```

RESULT 6

A40201  
artifact-warning sequence (translated ALU class A) - human  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000  
C:Accession: A40201  
R:Claverie, J.M.  
personal communication, 1992  
A:Reference number: A40201  
A:Accession: A40201  
A:Molecule type: DNA

A:Residues: 1-627 <CLA>

R:Claverie, J.M.  
Genomics 12, 838-841, 1992  
A:Title: Identifying coding exons by similarity search: Alu-derived and other potenti  
A:Reference number: A40200; MUID:92241891; PMID:1572661  
A:Contents: annotation  
C:Comment: this "warning" entry is a conceptual translation in all 6 reading frames o  
in-frame stop codons are shown as 'X'.  
C:Comment: Any significant similarity of a predicted protein sequence to a portion of

Alignment Scores:  
Pred. No.: 2.2e-23 Length: 627  
Score: 329.50 Matches: 90  
Percent Similarity: 67.52% Conservative: 16  
Best Local Similarity: 57.32% Mismatches: 35  
Query Match: 6.66% Indels: 17  
DB: 4 Gaps: 3

US-09-966-880A-7 (1-2818) x A40201 (1-627)

```
QY 1715 CTGCTTGGAGGCTTGGAGGAGGATGCTTGAACACAGGTGT----- 1758
Db 56 LeuLeuGlyArgLeuArgGlnGluAsnGlyValAsnProGlyGlyGlyAlaCysSerGlu 75
QY 1759 -----TCAAGG-----CCAGCCTGGGCAACATACAAGATCCTCTCTCAAAAA 1803
Db 76 ProArgSerArgHisCysThrProAlaLeuAlaThrGluArgAspSerValSerGluLys 95
QY 1804 AAAAAAAAAAAGAAAGA---GAGAGGCGCGCGGTGGTGGCTCAGCGCTGTANTCC 1860
Db 96 AsnLysAsnLysLys***Lys*****AlaGlyArgGly--LeuThrProValIleP 115
QY 1861 CAGCAGCTTTGGAGGCGCGGCGGATCACCTGTGTCAGGAGTTTGACACAGCGCT 1920
Db 115 roAlaLeuTrpGluAlaLysAlaGlyGlySerArg--GlyGlnGluIleGluThrIleLe 134
QY 1921 GCCCAACATGGCAAAACCCCGTCTGTACTCAAAATGCAAAATTAGCCAGGCGTGGTAGC 1980
Db 134 uAlaThrThrValLysProArgLeuTyr***LysTyrLysLysLeuAlaGlyArgSerAl 154
QY 1981 AGGCACCTCTAATCCAGCTACTTGGGCGGTGAGCAGGAGAATCGCTTGAACCCAGCA 2040
Db 154 aGlyAlaCysSerProSerTyrLeuGlyGly***GlyArgArgMetAla**ThrArgG 174
QY 2041 GGTGAGGTTCAGTAAGCTGAGATCGTCCCTTGCACCTCCAGCTGGCGCAACAGCA 2100
Db 174 uAlaGluLeuAlaValSerArgAspArgAlaThrAlaLeuGlnProTrpArgGln-SerG 194
QY 2101 AGACTCTGCTCAGAAAAA*****AAGAGAGAGAGAG 2141
Db 194 LuThrProSerArgLysLysLysLysLysThrLysAsnLysLys 207
```

RESULT 7

B40201  
artifact-warning sequence (translated ALU class B) - human  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000  
C:Accession: B40201  
R:Claverie, J.M.  
personal communication, 1992  
A:Reference number: A40201  
A:Accession: B40201  
A:Molecule type: DNA  
A:Residues: 1-301 <CLA>  
R:Claverie, J.M.  
Genomics 12, 838-841, 1992  
A:Title: Identifying coding exons by similarity search: Alu-derived and other potenti  
A:Reference number: A40200; MUID:92241891; PMID:1572661  
A:Contents: annotation  
C:Comment: This "warning" entry is a conceptual translation in all 6 reading frames o  
in-frame stop codons are shown as 'X'.  
C:Comment: Any significant similarity of a predicted protein sequence to a portion of







A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-116 <NAD>

A:Cross-references: EMBL:U03891; NID:9436940; PIDN:AAA03706.1; PID:9436941

C:Superfamily: apolipoprotein B mRNA editing enzyme, catalytic chain 1

Alignment Scores:  
Pred. No.: 2.98e-15 Length: 116  
Score: 245.00 Matches: 51  
Percent Similarity: 61.82% Conservative: 17  
Best Local Similarity: 46.36% Mismatches: 32  
Query Match: 4.95% Indels: 10  
DB: 2 Gaps: 3

US-09-966-880A-7 (1-2818) x G01233 (1-116)

QY 305 TACCGCGTCACTGGTTACCTCTGAGCGCCCTGTACGAC-----TGTGCCCCACAT 358  
|||||  
Db 7 TyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGlyGlu 26  
QY 359 GTGGCGGACTTCTGGAGGACCCCAACCTCAGTCTGAGGATCTTCACCGCGCCCTC 418  
|||||  
Db 27 ValArgAlaPheLeuGlnGlnThrHisValArgLeuProIlePheAlaAlaArgIle 46  
QY 419 TACTTCTGTGAGGACCGCAAGGCTGAGCC-----GAGGGGCTGGCGGCTGCAC 469  
|||||  
Db 47 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 61  
QY 470 CGCGCGGGGTGCAATATGCCATCATGACCTCAAGATATTTTACGTGCGGAATACT 529  
|||||  
Db 62 AspAlaGlyAlaGlnValSerIleMetThrTrpAspGluPheGluTyrCysTrpAspThr 81  
QY 530 TTGTGTAGAAACCATGAAAGAACTTCAAGCTGGGAAGGCTGCATGAAATTCAGTT 589  
|||||  
Db 82 PheValTyrArgGlnGlyCysProPheGlnProTrpAspGlyLeuGluGlnHisSerGln 101  
QY 590 CGTCTCCAGACAGCTTCGGCGCAATCCTT 619  
|||||  
Db 102 AlaLeuSerGlyArgLeuArgAlaIleLeu 111

RESULT 13

A46010

X-linked retinopathy protein (C-terminal, clone XEH.8c) - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 05-Nov-1999

C:Accession: A46010

R:Wong, P.; MacDonald, I.M.; Sood, R.; Smith, C.; Pilon, R.; Tenniswood, M.

Genomics 15, 467-471, 1993

A:Title: Identification and partial characterization of a candidate gene for X-linked re

A:Reference number: A46010; MUID:93224131; PMID:8468040

A:Accession: A46010

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-100 <WON>

A:Cross-references: GB:S58722; NID:9299470; PIDN:AAB26149.1; PID:9299471

A:Note: sequence extracted from NCBI backbone (NCBIN:129339, NCBI:129340)

Alignment Scores:  
Pred. No.: 4.55e-15 Length: 100  
Score: 243.00 Matches: 63  
Percent Similarity: 65.18% Conservative: 10  
Best Local Similarity: 56.25% Mismatches: 27  
Query Match: 4.92% Indels: 12  
DB: 2 Gaps: 3

US-09-966-880A-7 (1-2818) x A46010 (1-100)

QY 2130 TTTTNTTTTCTTGAGACAGAGTCTTGCTTGTGCGCCAGGCTGGAGTGCAACG 2071

|||||

Db 1 PhePhePhePheGluThrGluSerCysSer-ValAlaGluAlaGlyValGlnTr 20

QY 2070 GCACGATCTCACTTACTGCAACCTCCACCTCCTGGGTTCAACGCGATTCCTCGCTCAG 2011

|||||

|||||

Db 20 pCysAspLeuGlySerLeuLysSerProProGlySerSerAspSerProAlaSerA 40  
QY 2010 CCTCCCAAGTAGCTGGATTACAGGTCCCTGCTACACGCGCTGGTAATTTTGCATTTT 1951  
|||||  
Db 40 lAserArgValAlaGlyIleThrGlyMetHisHisThrGlnLeuIlePheValPhe 60  
QY 1950 GAGTACAGACGGGTTTTCATGTTGGCCAGCTGGTCTCAAACTCCTCACCACAGGTG 1891  
|||||  
Db 60 euValGluThrGly---SerHisMetGlnLeuSerAspSerThrLeuValIleThr---- 77  
QY 1890 ATCCGCCGCGCTCGGCTCCCAAGTGTGGGATTACAGCGGTGAGCCACCGCCCGGC 1831  
|||||  
Db 78 -----ThrAlaGlnAsnAlaLysIleThrAla-----ArgAlaProA 90  
QY 1830 CCTCTCTCTTCTTTTCTTTTCTTTTCTTTT 1799  
|||||

Db 90 rGAspLeuPhePhePhePhePhePhePhe 100

RESULT 14

A53853

apolipoprotein B mRNA editing enzyme, catalytic chain 1 (EC 3.5.4.-) - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 18-Aug-2000

C:Accession: A53853

R:Yamanaka, S.; Poksay, K.S.; Balestra, M.E.; Zeng, G.Q.; Innerarity, T.L.

J. Biol. Chem. 269, 21725-21734, 1994

A:Title: Cloning and mutagenesis of the rabbit ApoB mRNA editing protein. A zinc mot

ributed.

A:Reference number: A53853; MUID:94342367; PMID:8063816

A:Accession: A53853

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-236 <YAM>

A:Cross-references: GB:U10695; NID:g506180; PIDN:AAA56718.1; PID:g506181

C:Superfamily: apolipoprotein B mRNA editing enzyme, catalytic chain 1

C:Keywords: hydrolase; zinc

Alignment Scores:

Pred. No.: 3.12e-13 Length: 236  
Score: 225.00 Matches: 49  
Percent Similarity: 57.25% Conservative: 26  
Best Local Similarity: 37.40% Mismatches: 48  
Query Match: 4.55% Indels: 8  
DB: 2 Gaps: 4

US-09-966-880A-7 (1-2818) x A53853 (1-236)

QY 149 CGCGGTGAGACCTACCTGTCTACGTAGTGAAGAGCGGTGACAGTCTACATCTTTTCA 208

|||||

Db 33 ArgLysGluAlaCysLeuLeuTyrGluIleLysTrpGlyAlaSerSerLysThrTrpArg 52

QY 209 CTGGACTTTGGTTATCTTCGCAATAGACGGCTGC---CACGTGAANTTGTCTTCCTC 265

|||||

Db 53 SerSer-----GlyLysAsnThrThrAsnHisValGluValAsnPheLeu 67

QY 266 ---CGGTACATCTCGGACTGGGACCTAGACCTCGGCGGTGACCGGTACCTGGTTC 322

|||||

Db 68 GluLysLeuThrSerGluGlyArgLeuGlyProSerThrCysCysSerIleThrTrpPhe 87

QY 323 ACTCTCTGAGGCCCTGTCTACGACTGTGCCGACATGTGGCCGACTTTCTGCGAGGAAC 382

|||||

Db 88 LeuSerTrpSerProCysTrpGluCysSerMetAlaIleArgGluPheLeuSerGlnHis 107

QY 383 CCCAACTCAGTCTGAGGATCTTCACCGCGGCTCTACTTCTGTGAGGACCCCAAGGCT 442

|||||

Db 108 ProGlyValThrLeuIleIlePheValAlaArgLeuPheGlnHisMetAspArg--- 126

QY 443 GAGCCCGAGGGCTGCGGGCTGACCGCGCGGGTGCACAAATAGCCATCATGACCTTC 502

|||||

Db 127 AsnArgGlnGlyLeuLysAspLeuValThrSerGlyValThrValArgValMetSerVal 146

QY 503 AAAGATTATTTTACTGCTGGAATACTTTTGTGA 535

|||||

```

Db      147 SerGluTyrCysTyrCysTrpGluAsnPheVal 157
RESULT 15
S41044
chromosomal protein - human
C:Species: Homo sapiens (man)
C:Date: 19-May-1994 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000
C:Accession: S41044
R:Yeo, J.P.; Alderuccio, F.; Toh, B.H.
Nature 367, 288-291, 1994
A:Title: A new chromosomal protein essential for mitotic spindle assembly.
A:Reference number: S41044; MUID:94166884; PMID:8121495
A:Accession: S41044
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-418 <YEO>
A:Cross-references: GB:L26953; NID:g537529; PIDN:AAB68050.1; PID:g537530
C:Superfamily: human 48.2K chromosomal protein
C:Keywords: Chromosomal protein

Alignment Scores:
Pred. No.:      4.9e-13      Length:      418
Score:          223.50      Matches:      64
Percent Similarity: 51.61%      Conservative: 32
Best Local Similarity: 34.41%      Mismatches: 36
Query Match:      4.52%      Indels:      54
DB:              2          Gaps:      7

US-09-966-880A-7 (1-2818) x S41044 (1-418)

QY      1794 CTCTCAAAAAAAAAAAAAA----- 1814
Db      127 IleSerGlnLysLysLysArgGlyIleTyrHisLysAsnAsnIleArgIleIleLeuPhe 146
QY      1815 AAAAGAAGAGAGAGCGCGCGGTGGTGGCTGACGCTGTAATCCCA----- 1862
Db      147 LeuProGlnAlaHisGlyArgAspPheTyrVal---ProIleLeuProPheThrGlnSer 165
QY      1863 -----GCACCTTTGGAGCGCGCGCGGATCACCTGTGGTC 1901
Db      166 TyrValAspTTPGlyArgTrpLeuIleTrpGluAlaLysAlaGlyGluSerLeuGluVal 185
QY      1902 AGGAGTTTGACACAGCGCTGCCCAACATGCCAAAC----- 1937
Db      186 ArgSerArgProAlaSerGlnSerArgArgAsnSerValSerThrLysAsnIleLys 205
QY      1938 -----CCCGTCTGTACTCAAAATGCAAAATTAGCCAGCGCTGGTAGCAGGCACCT--- 1988
Db      206 IleSerProValSerThrLysAsnIleLysIleSerGlnThrTrpTyrLeuPheGlyGly 225
QY      1989 -----GTAATCCAGCTACTTGGAGGCTGAGCGAGAGAAATCGCTTGAACCCAGG 2039
Db      226 ValHisLeuLeuValProThrThrArgAspAlaGluAlaGlyGluLeuHisAsp-ProG1 245
QY      2040 AGGTGGAGGTTCAGTAAGCTGAGATCGTCCCTTCCTCCTCAGCTCCAGCTGGCGCAGAGAGC 2099
Db      245 YGlyArgGlyCysAsnGluLeuArgSerCysHisCysThrProAlaTrpValThr--Ser 264
QY      2100 AAGACTCTGTCTCAGAAAAA----- 2147
Db      265 GluThrValSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 284
QY      2148 -----AGACAAATATTGGGAGAGAGGATGGGAAGCATTCGAAGGAAATGTGC 2198
Db      285 IleAsnAlaSerThrLeuPhe-----HisValLeuThrArgIlePheCys 299
QY      2199 TTTATCCAAACA 2210
Db      300 TyrLysGlnLys 303

```

Search completed: June 14, 2003, 18:56:28  
Job time : 98.1094 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 14, 2003, 18:11:13 ; Search time 35.3473 Seconds  
(without alignments)  
6613.252 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 4948  
Sequence: 1 agagaacatcatcattatga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/sgn2.1/USPTO-spool/US09966880/runat\_14062003\_175524\_10304/app\_query.fasta\_1.9493  
-DB=SwissProt\_40 -Qfmt=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09966880 -ECGN\_1\_1\_196 -runat\_14062003\_175524\_10304 -NCPU=6 -ICPU=3  
-NO\_WMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES						
Result No.	Score	Match	Length	ID	Description	
C 1	498	10.1	593	1	ALU7_HUMAN	P39194 homo sapien
C 2	484	9.8	593	1	ALU6_HUMAN	P39193 homo sapien
C 3	457	9.2	593	1	ALU7_HUMAN	P39194 homo sapien
C 4	430	8.7	593	1	ALU6_HUMAN	P39193 homo sapien
C 5	420.5	8.5	591	1	ALU1_HUMAN	P39188 homo sapien
C 6	416.5	8.4	591	1	ALU1_HUMAN	P39188 homo sapien
C 7	414	8.4	591	1	ALU8_HUMAN	P39195 homo sapien
C 8	382.5	7.7	591	1	ALU8_HUMAN	P39195 homo sapien
C 9	372	7.5	585	1	ALU5_HUMAN	P39192 homo sapien
C 10	368.5	7.5	587	1	ALU2_HUMAN	P39189 homo sapien
C 11	366	7.4	585	1	ALU5_HUMAN	P39192 homo sapien
C 12	364	7.4	382	1	PHB3_HUMAN	Q9uh17 homo sapien
C 13	363.5	7.3	199	1	PHB1_HUMAN	P31941 homo sapien
C 14	359.5	7.3	587	1	ALU3_HUMAN	P39190 homo sapien
C 15	358	7.2	587	1	ALU3_HUMAN	P39190 homo sapien
C 16	356	7.2	587	1	ALU2_HUMAN	P39189 homo sapien
C 17	339	6.9	190	1	PHB2_HUMAN	Q9ue74 homo sapien
C 18	295	6.0	603	1	ALU4_HUMAN	P39191 homo sapien

19	288.5	5.8	603	1	ALU4_HUMAN	P39191 homo sapien
20	225	4.5	236	1	ABME_RABIT	P47855 oryctolagus
21	223.5	4.5	418	1	YY1_HUMAN	P49646 homo sapien
22	221	4.5	236	1	ABME_HUMAN	P41238 homo sapien
23	213.5	4.3	229	1	ABME_MOUSE	P51908 mus musculus
C 24	199.5	4.0	229	1	ABME_RAT	P38483 rattus norv
C 25	182.5	3.7	841	1	NEK4_HUMAN	P51957 homo sapien
C 26	143.5	2.9	124	1	YY3_HUMAN	P20931 homo sapien
C 27	134	2.7	447	1	KBF3_HUMAN	Q04860 homo sapien
C 28	130	2.6	369	1	TA2R_HUMAN	P21731 homo sapien
C 29	128.5	2.6	881	1	PRP2_HUMAN	Q99599 homo sapien
C 30	120	2.4	619	1	REL_HUMAN	Q04864 homo sapien
C 31	119	2.4	629	1	Z195_HUMAN	Q14628 homo sapien
C 32	108	2.2	1191	1	ZN91_HUMAN	Q05481 homo sapien
C 33	106	2.1	1371	1	UBP1_HUMAN	Q04966 homo sapien
C 34	98.5	2.0	369	1	CT19_YEAST	Q02732 saccharomyc
C 35	94	1.9	504	1	AT1N_HSVBP	P30020 bovine herp
C 36	92	1.9	2476	1	ATRX_MOUSE	Q61687 mus musculu
C 37	91	1.8	1402	1	IF4G_RABIT	P41110 oryctolagus
C 38	90.5	1.8	522	1	NU62_HUMAN	P37198 homo sapien
C 39	89.5	1.8	976	1	BUL1_YEAST	P48524 saccharomyc
C 40	89	1.8	354	1	WN14_CHICK	O42280 gallus gall
C 41	89	1.8	634	1	YCK3_EUGER	P31916 euglena gra
C 42	89	1.8	1064	1	ISK5_HUMAN	Q9nq38 homo sapien
C 43	88.5	1.8	619	1	VAL1_YEAST	P38085 saccharomyc
C 44	88	1.8	735	1	PSAB_CHLRE	P09144 chlamydomon
C 45	88	1.8	914	1	GNDS_HUMAN	Q12967 homo sapien

ALIGNMENTS

RESULT 1  
ALU7\_HUMAN  
ID ALU7\_HUMAN STANDARD; PRT; 593 AA.  
AC P39194;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alu subfamily S0 sequence contamination warning entry.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95021758; PubMed=7935834;  
RA Claverie J.-M., Makalowski W.;  
RT "Alu alert.";  
RL Nature 371:752-752(1994).  
RN [2]  
RP CONCEPT.  
RX MEDLINE=92241891; PubMed=1572661;  
RA Claverie J.-M.;  
RT "Identifying coding exons by similarity search: alu-derived and other  
potentially misleading protein sequences.";  
RL Genomics 12:838-841(1992).  
RN [3]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=88333009; PubMed=3138422;  
RA Quentin Y.;  
RT "The Alu family developed through successive waves of fixation  
closely connected with primate lineage history.";  
RL J. Mol. Evol. 27:194-202(1988).  
RN [4]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=91178815; PubMed=1706781;  
RA Jurka J., Milosavljevic A.;  
RT "Reconstruction and analysis of human Alu genes.";  
RL J. Mol. Evol. 32:105-121(1991).  
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU





CC	CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE	
CC	OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A	
CC	CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING	
CC	DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH	
CC	THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES	
CC	BEING REPORTED.	
CC	-1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE	
CC	WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A	
CC	PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE	
CC	CODING NUCLEOTIDE SEQUENCE.	
CC	-----	
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
CC	EMBL; U14573; ; NOT_ANNOTATED_CDS.	
KR	Hypothetical protein.	
KW	DOMAIN 1 97	
FT	FRAME-1.	
FT	DOMAIN 101 196	
FT	FRAME-2.	
FT	DOMAIN 200 295	
FT	FRAME-3.	
FT	DOMAIN 299 395	
FT	FRAME-4.	
FT	DOMAIN 399 494	
FT	FRAME-5.	
FT	DOMAIN 498 593	
FT	FRAME-6.	
SQ	SEQUENCE 593 AA: 64417 MW: 54A4F50F33A6089F CRC64;	
CC	-----	
CC	Alignment Scores:	
Pred. No.:	9.94e-33	Length: 593
Score:	457.00	Matches: 98
Percent Similarity:	56.77%	Conservative: 11
Best Local Similarity:	51.04%	Mismatches: 33
Query Match:	9.24%	Indels: 50
DB:	1	Gaps: 1
US-09-966-880A-7 (1-2818) x ALU7_HUMAN (1-593)		
QY	1693 TGGCACAGCTCATGTTCTAGCTGTGGAGGTTGAGGAGGAGGATGCTTCAACAC	1752
DB	5 TrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlySerProGluVal	24
QY	1753 AGGTCTTCAAGCCAGCTGGCGACATAACAAGATCTGCTCTCAAAAAA	1812
DB	25 ArgSerSerArgProAlaTrpProThrTrp***AsnProValSerThrLysAsnThrLys	44
QY	1812 -----	1812
DB	45 IleSerArgAlaTrpTrpArgAlaProValIleProAlaThrArgGluAlaGly	64
QY	1812 -----	1812
DB	65 GluSerLeuGluProGlyArgArgArgLeuGln***AlaGluIleAlaProLeuHisSer	84
QY	1813 -----	1813
DB	85 SerLeuGlyAsnLysSerGluThrProSerGlnLysLys*****AlaGlyArgGly	104
QY	1843 GCCTACGCTGTATCCAGCACATTGGAGCGCGAGCGCGGATACCTGTGGTCA	1902
DB	105 GlySerArgLeu***SerGlnHisPheGlyArgProArgArgValAspHisLeuArgSer	124
QY	1903 GGAGTTTGACGAGCCTGGCCCAACATGGCAAAACCCGCTCTGTACTCAAAATGCAAAA	1962
DB	125 GlyValArgAspGlnProGlyGlnHisGlyGluThrProSerLeuLeuLysIleGlnLys	144
QY	1963 TTAGCCAGCGGTGGTAGCAGGACCTGTATCCAGCTACTTGGGAGGCTGAGGAGGAG	2022
DB	145 LeuAlaGlyArgGlyGlyArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGlu	164
QY	2023 AATCGCTTCAACCCAGGAGGTGGAGGTTGCAGTAAAGCTGAGATCGTGGCGGTGCACCTCA	2082

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL; U14572; -; NOT\_ANNOTATED\_CDS.

KW Hypothetical protein.

FT DOMAIN 1 97 FRAME-1.  
FT DOMAIN 101 196 FRAME-2.  
FT DOMAIN 200 295 FRAME-3.  
FT DOMAIN 299 395 FRAME-4.  
FT DOMAIN 399 494 FRAME-5.  
FT DOMAIN 498 593 FRAME-6.  
SQ SEQUENCE 593 AA; 64603 MW; 136EF344AACD12A2 CRC64;

#### Alignment Scores:

Pred. No.: 2.6e-30 Length: 593  
Score: 430.00 Matches: 95  
Percent Similarity: 55.73% Conservative: 12  
Best Local Similarity: 49.48% Mismatches: 35  
Query Match: 8.69% Indels: 50  
DB: 1 Gaps: 1

US-09-966-880A-7 (1-2818) x ALU6\_HUMAN (1-593)

QY 1693 TGGCACCGCTATAGTTCTAGCTGCTGGAGGTGAGGAGGAGGATGGCTTGAACAC 1752  
DB 5 TrpLeuThrProValIleProAlaLeuTrpGluAlaGluAlaGlySerProGluVal 24  
QY 1753 AGGTGTGAAGCCAGCGTGGCCACATCAACAGATCCCTCTCTCAAAAAA 1812  
DB 25 GlySerSerArgProAla\*\*\*ProThrTrpArgAsnProValSerThrLysAsnThrLys 44  
QY 1812 ----- 1812  
DB 45 IleSerArgAlaTrpTrpArgMetProValIleProAlaThrArgGluAlaGluAlaGly 64  
QY 1812 ----- 1812  
DB 65 GluSerLeuGluProGlyArgArgArgLeuArg\*\*\*AlaGluIleAlaProLeuHisSer 84  
QY 1813 -----AAAAAGAAAGAGAGAGGCGCGGTGCT 1842  
DB 85 SerLeuGlyAsnLysSerGluThrProSerGlnLysLys\*\*\*\*\*AlaGlyArgGly 104  
QY 1843 GGCTCAGCCTGTATCCAGACACTTGGAGCGCGGAGCGCGGATCACCTGTGGTCA 1902  
DB 105 GlySerArgLeu\*\*\*SerGlnHisPheGlyArgProArgArgAlaAspHisLeuArgSer 124  
QY 1903 GGAGTTTGACAGCAGCTGGCCCAACATGCAAAACCCGCTGTACTCAAAATGCAAAA 1962  
DB 125 GlyValArgAspGlnProAspGlnHisGlyGluThrProSerLeuLeuLysIleGlnLys 144  
QY 1963 TTAGCCAGCGGTGGTAGCAGGACCTGTATCCAGCTACTTGGAGCGCTGAGCGAGGAG 2022  
DB 145 LeuAlaGlyArgGlyGlyAlaCysLeu\*\*\*SerGlnLeuLeuGlyArgLeuArgGlnGlu 164  
QY 2023 ATTCGCTGAACCCAGGAGGTGGAGTTGACGTAGCTAGCTGAGTGGCGGTGGCATCCA 2082  
DB 165 AsnArgLeuAsnProGlyGlyGlyGlyCysGlyGluProArgSerArgHisCysThrPro 184  
QY 2083 GCCTGGCCACAGACAGACTCTCTCAGAAA 2118  
DB 185 AlaTrpAlaThrArgAlaLysLeuArgLeuLysLys 196

#### RESULT 5

ALU1\_HUMAN  
ID ALU1\_HUMAN STANDARD; PRT; 591 AA.  
AC P39188;  
DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alu subfamily J sequence contamination warning entry.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RN SEQUENCE FROM N.A.  
RX MEDLINE=95021758; PubMed=7935834;  
RA Claverie J.-M., Makalowski W.;  
RT "Alu alert.";  
RL Nature 371:752-752(1994).  
RN [2]  
RN CONCEPT.  
RX MEDLINE=92241891; PubMed=1572661;  
RA Claverie J.-M.;  
RT "Identifying coding exons by similarity search: alu-derived and other  
RT potentially misleading protein sequences.";  
RL Genomics 12:838-841(1992).  
RN [3]  
RN ALU FAMILIES CLASSIFICATION  
RX MEDLINE=88333009; PubMed=3138422;  
RA Quentin Y.;  
RT "The Alu family developed through successive waves of fixation  
RT closely connected with primate lineage history.";  
RL J. Mol. Evol. 27:194-202(1988).  
RN [4]  
RN ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=91178815; PubMed=1706781;  
RA Jurka J., Milosavljevic A.;  
RT "Reconstruction and analysis of human Alu genes.";  
RL J. Mol. Evol. 32:105-121(1991).  
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
CC REPEATS.  
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
CC CODON. 'XX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
CC ACID SEQUENCES.  
CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
CC BEING REPORTED.  
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
CC CODING NUCLEOTIDE SEQUENCE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U14567; -; NOT\_ANNOTATED\_CDS.  
KW Hypothetical protein.  
FT DOMAIN 1 96 FRAME-1.



FT	DOMAIN	100	195	FRAME-2.
FT	DOMAIN	199	294	FRAME-3.
FT	DOMAIN	298	393	FRAME-4.
FT	DOMAIN	397	492	FRAME-5.
FT	DOMAIN	496	591	FRAME-6.
SQ	SEQUENCE	591 AA;	63790 MW;	665D395735519D95 CRC64;

  

Alignment Scores:	
Pred. No.:	1.84e-29
Score:	420.50
Percent Similarity:	43.88%
Best Local Similarity:	39.46%
Query Match:	8.52%
DB:	1

  

US-09-966-880A-7 (1-2818) x ALU1_HUMAN (1-591)	
Qy	2359 TCCTGCTCATCTTCATTCGCAAGGTGCTAGGTAATAACGGCACAGGATATGGTTATTCT 2300
Db	307 SerProArgLeuGluCysSerGly-----AlaIleThrAla----- 318
Qy	2299 GAATATTTTCTCAGCTGCTCAAAAGACAGTGTGTAGGCACAAT- 2258
Db	319 -----HisCysSerLeuAspLeuProGlySerSerAspProProAlaSerAla 334
Qy	2257 -----AGACACCAAAAGACAAATAGGATGGATCCCTTAT 2225
Db	335 SerArgValAlaGlyThrThrGlyAlaArgHisAlaArgLeuIlePheValPhePhe 354
Qy	2224 TGGCTCCTTACATTTTGTGGTAAGCAACAATTCCTTGCAGTCTCCCATCCTTCT 2165
Db	355 ValGluThrGlyPheHisIstyrValAlaGlnAlaGlyLeuGluLeuGlySerSerAsp 374
Qy	2164 CTCCCA-----AATATTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTTT 2126
Db	375 ProProAlaSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArg*** 394
Qy	2125 TTTTNTTTTTTCTGAGCAGAGTCTTCTCTGTGCGCCAGGCTGGAGTGCAAGCGCAGC 2066
Db	395 *****PhePhe***AspArgValSerLeu-CysArgProGlyTrpSerAlaValAlaAr 414
Qy	2065 ATCTCAGCTTACTGCAACCTCCACTCTCTGGGTTCAGCGATCTCTCTGCTCAGCTCC 2006
Db	414 gSerArgLeuThrAlaAlaSerThrSerArgAlaGlnAlaIleLeuLeuProGlnProPr 434
Qy	2005 CAAGTAGCTGGATTACAGTGCGCTGCTACCAGCCTGGCTAATTTTGCATTTTGAGTA 1946
Db	434 oGlu***LeuGlyLeuGlnAlaArgAlaThrThrProGly***PheLeuTyrPheLeu** 454
Qy	1945 CAGACGGGGTTTGGCAATGTGGCCAGGCTGCTCAAACTCCTGACCACAGGTGATCCG 1886
Db	454 *ArgArgGlyPheThrMetLeuProArgLeuValSerAsnSerTrpAlaGlnValIleLe 474
Qy	1885 CCGGGTCGCGCTCCCAAAGTGTGGATTACAGGCGTGAGCCACCGCCCGCCCTCT 1826
Db	474 uProProArgProProLysValLeuGlyLeuGlnAla***AlaThrAlaProGly**** 494
Qy	1825 CTCTCTCTTT----- 1816
Db	494 ***PhePheGluThrGlySerArgSerValAlaGlnAlaGlyValGlnTrpArgAspHI 514
Qy	1816 ----- 1816
Db	514 sGlySerLeuGlnProArgProProGlyLeuLysArgSerSerCysLeuSerLeuProSe 534
Qy	1815 -----TTTTTTTTTTTTTTTTTTTGAGAGA 1793
Db	534 rSerTrpAspTyrArgArgAlaProProArgProAlaAsnPheCysIlePheCysArgAs 554
Qy	1792 CAGGATCTTGTTATCTTGCACGAGCTGCCCTTGCACACCTGTGTCAAGCCATCTCCCT 1733
Db	554 pGlyValSerLeuCysCysProGlyTrpSerArgThrProGlyLeuLys***SerSerAr 574







CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
CC BEING REPORTED.  
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
CC CODING NUCLEOTIDE SEQUENCE.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL: U14574; -; NOT\_ANNOTATED\_CDS.  
CC KW Hypothetical protein.  
CC FT DOMAIN 1 96 FRAME-1.  
CC 100 195 FRAME-2.  
CC 199 294 FRAME-3.  
CC 298 393 FRAME-4.  
CC 397 492 FRAME-5.  
CC 496 591 FRAME-6.  
CC SQ SEQUENCE 591 AA; 64395 MW; AC8154AD8A6B280 CRC64;

Alignment Scores:  
Pred. No.: 4, 678-26 Length: 591  
Score: 382.50 Matches: 103  
Percent Similarity: 68.12% Conservative: 6  
Best Local Similarity: 64.38% Mismatches: 32  
Query Match: 7.73% Indels: 20  
DB: 1 Gaps: 5

US-09-966-880A-7 (1-2818) x ALU8\_HUMAN (1-591)

QY 1679 AAGCTGCGAACCGT---GGCACAGCTCATAGTTCTAGCTGCTGGAGGTGAGGAG 1735  
Db 143 LysLeuAlaGlyArgGlyAlaArgLeu\*\*\*SerGlnLeuLeuGlyArgLeuArgGln 162  
QY 1736 GAGGATGCTTCACACAGGTGT-----TCAAGG----- 1764  
Db 163 GluAsnArgLeuAsnProGlyGlyGlyCysSerGluProArgSerArgHisCysThr 182  
QY 1765 CCAGCTGGGCAACATAACAGATCTCTCTCAAAAAAATAAAGAAAGAAAGA 1824  
Db 183 ProAlaTrpAlaThrGluArgAspSerValSer-----LysLys\*\*\*\*\*Arg 199  
QY 1825 GAGGGCGGGCGTGGTGGCTGCTGTAATCCAGCACTTTGGAGGCCGAGCCGG 1884  
Db 200 AlaGluAlaGlyArgGlySerArgLeu\*\*\*SerGlnHisPheGly-----Arg 216  
QY 1885 CGGGATCACCTGTGTCAGGAGTTTGACACAGCCCTGGCCACATGGCAAAACCCGCT 1944  
Db 217 -LysileThr\*\*\*GlyGlnGluPheGluThrSerLeuAlaAsnMetValLysProArgLe 236  
QY 1945 GTACTCAAAATGCAAAATATTAGCCAGCGTGGTAGCAGCCTCTAATCCAGCTACTT 2004  
Db 236 uTyR\*\*\*LysTyRAsn\*\*\*ProGlyValValAlaArgAlaCysAsnProSerTySe 256  
QY 2005 GGGAGGCTGAGCAGGAGAGTGGTGAACCCAGGAGGTGGAGGTGAGTGAAGCTGAGA 2064  
Db 256 rGlyGly\*\*\*GlyArgArgIleAla\*\*\*ThrArgGluAlaGluValAlaValSerArgAs 276  
QY 2065 TCGTCCCTGCTGACTCCAGCTGGCGGACAGACAGCAACTGCTCTCAGAAAAAA 2120  
Db 276 pArgAlaThrAlaLeuGlnProGlyArgGln-SerGluThrProSerGlnLysLys 294

RESULT 9  
ALU5\_HUMAN  
ID ALU5\_HUMAN STANDARD: PRT; 585 AA.  
AC P39192;  
DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Alu subfamily SC sequence contamination warning entry.  
OS Homo sapiens (Human)  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95021758; PubMed=7935834;  
RA Claverie J.-M., Makalowski W.;  
RT "Alu alert.";  
RT Nature 371:752-752(1994).  
RN [2]  
RP CONCEPT.  
RX MEDLINE=92241891; PubMed=1572661;  
RA Claverie J.-M.;  
RT "Identifying coding exons by similarity search: alu-derived and other  
RT potentially misleading protein sequences.";  
RL Genomics 12:838-841(1992).  
RN [3]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=88333009; PubMed=3138422;  
RA Quentin Y.;  
RT "The Alu family developed through successive waves of fixation  
RT closely connected with primate lineage history.";  
RL J. Mol. Evol. 27:194-202(1988).  
RN [4]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=91178815; PubMed=1706781;  
RA Jurka J., Milosavljevic A.;  
RT "Reconstruction and analysis of human Alu genes.";  
RL J. Mol. Evol. 32:105-121(1991).  
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE  
CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
CC REPEATS.  
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
CC CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
CC ACID SEQUENCES.  
CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
CC CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
CC BEING REPORTED.  
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
CC CODING NUCLEOTIDE SEQUENCE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL: U14571; -; NOT\_ANNOTATED\_CDS.  
KW Hypothetical protein.  
FT DOMAIN 1 95 FRAME-1.

[1] SEQUENCE FROM N.A.  
RX MEDLINE=95021758; PubMed=7935834;  
RA Claverie J.-M., Makalowski W.;  
RT "Alu alert.";  
RL Nature 371:752-752(1994).  
RN [2]  
RP CONCEPT.  
RX MEDLINE=92241891; PubMed=1572661;  
RA Claverie J.-M.;  
RT "Identifying coding exons by similarity search: alu-derived and other  
potentially misleading protein sequences.";  
RL Genomics 12:838-841(1992).  
RN [3]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=88333009; PubMed=3138422;  
RA Quentin Y.;  
RT "The Alu family developed through successive waves of fixation  
closely connected with primate lineage history.";  
RL J. Mol. Evol. 27:194-202(1988).  
RN [4]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=91178815; PubMed=1706781;  
RA Jurka J., Milosavljevic A.;  
RT "Reconstruction and analysis of human Alu genes.";  
RL J. Mol. Evol. 32:105-121(1991).  
CC -!- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
REPEATS.  
CC -!- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
CODON, 'xxx' IS USED TO INDICATE THE VARIOUS TRANSLATION PHASES.  
CC -!- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER  
POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO  
ACID SEQUENCES.  
CC -!- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND  
PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE  
ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN  
ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,  
CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS  
LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU  
ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A  
GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,  
CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE  
OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A  
CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING  
DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH  
THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES  
BEING REPORTED.  
CC -!- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE  
WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A  
PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE  
CODING NUCLEOTIDE SEQUENCE.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U14568; ; NOT\_ANNOTATED\_CDS.  
DR Hypothetical protein.  
KW 1 96  
FT DOMAIN 100 194 FRAME-1.  
FT FT 198 292 FRAME-2.  
FT DOMAIN 296 391 FRAME-3.  
FT DOMAIN 395 489 FRAME-4.  
FT DOMAIN 493 587 FRAME-5.  
FT DOMAIN 587 63703 MW. 3EAB9B3E26202021 6B0C64  
SQ SEQUENCE



Db	165	-----	GlupProGlyArgArgLeuGln***AlaGluIleAlaProLeuHi	180
		:		
QY	1783	CAAGATCTGTCTCTCAAAAAA	AAAAAAAAAAGAAAGAGAGAGGGCGGCGTGGT	1842
		:		
Db	180	sSerSerLeuAlaThrGluArgAspSerValSerLysLys*****ProGlyAlaVa	200	
		:		
QY	1843	GGCTCACCCCTGTAATCCAGCACTTTGGGAGCGCGAGCGGGCGGATACCTCTGTGTC	1902	
		:		
Db	200	lAlaHisAlaCysAsnProSerThrLeuGlyGlyArgGlyGlyArgIleThrArgSerAr	220	
		:		
QY	1903	GGAGTTTGAGACAGCCCTGGGCCAACATGGCAAAACCCCGTGTGTACTCAAAATGCAAAA	1962	
		:		
Db	220	gAspArg--AspHisProGlyGlnHisGlyGlyThrProSerLeuLeuIleGlnLys	239	
		:		
QY	1963	TTAGCCAGCGCTGTGTAGCAGCACCTCTAATCCAGCTACTTGGGAGCTTGAGGCAGGAG	2022	
		:		
Db	240	LeuAlaGlyArgGlyGlyAlaArgLeu***SerGlnLeuLeuGlyArgLeuArgGlnGlu	259	
		:		
QY	2023	AATCGCTTGAACCCAGGAGGTGGAGTTGCAGTAAGTGCATGCTGCCCTTGCATCTCCA	2082	
		:		
Db	260	AsnArgLeuAsnProGlyGlyGlyGlyCysSerGluProArgSerArgHisCysThrPro	279	
		:		
QY	2083	GCCTGGCGGACAGCAAGCACTGTCTCAGAAAAA	2142	
		:		
Db	280	AlaTrp-ArgGln-SerGluThrProSerGlnLysLys*****PhePheLeuArgA	299	
		:		
QY	2143	GAAGAGAACAAATATTTGGGAGAGAGGATGGGGAAGCATTTGCAAGG	2189	
		:		
Db	299	rgSerLeuAlaLeu-----SerProGlyTrpSerAlaValAlaArg	312	
		:		
RESULT 12				
PHB3_HUMAN				
ID	PHB3_HUMAN	STANDARD;	PRT;	382 AA.
AC	Q9UH17; Q95618;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Phorbolin 3 (APOBEC1-like).			
GN	APOBEC1L			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
NCBI	NCBI_TaxId=9606;			

## RESULT 12

ID	PHB3_HUMAN	STANDARD;	PRT;	382 AA.
AC	Q9UH17; Q95618;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Phorbol 3 (APOBEC1-like).			
GN	APOBEC1L.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
EN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20057165; PubMed=10591208;			
RA	Dunham I., Hunt A.R., Collins J.E., Bruskielwich R., Beare D.M.,			
RA	Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,			
RA	Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,			
RA	Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,			
RA	Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,			
RA	Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,			
RA	Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,			
RA	Dharm P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,			
RA	Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,			
RA	Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,			
RA	Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,			
RA	Hunt S.E., Jones M.C., Kerhaw J., Kimberley A.M., King A.,			
RA	Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,			
RA	Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., McCann O.T.,			
RA	Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,			
RA	Odeh C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,			
RA	Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,			
RA	Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,			
RA	Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,			
RA	Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,			
RA	Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,			
RA	Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,			
RA	Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,			
RA	Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyma S.,			
RA	Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,			
RA	Dorman A., Fang F., Fu Y., Hua A., Kenton S., Lai H., Lao H.I.,			

QY 356 CATGTGGCGGAGCTTTCTGCGAGGAGAACCCACCTCAGTCTGAGGATCTTCCGCCGCGC 415  
 DB 292 GluValArgAlaPheLeuGlnGluAsnThrHisValArgLeuArgilePheAlaLaArg 311  
 QY 416 CTCCTACTTCTGTGAGGACCGCAGGCTGAGCCC-----GAGGGGCTGGCGGCGCTG 466  
 DB 312 IleTyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeu 326  
 QY 467 CACCGCGCGGGTGCAATGCCATCATGACCTCAAGATATTTTACTGCTGGGAT 526  
 DB 327 ArgAspAlaGlyValGlnValSerIleMetThrTyrAspGluPheLeuTyrCysTrpasp 346  
 QY 527 ACTTTGTGAGAAACCATCAAGAACTTTCAAGCGCTGGGAGGCGTGCATGAAATTC 586  
 DB 347 ThrPheValTyrArgGlnGlyCysProPheGlnProTyrAspGlyLeuGluHisSer 366  
 QY 587 GTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 619  
 DB 367 GlnAlaLeuSerGlyArgLeuArgAlaIleLeu 377

RESULT 13  
 PHB1\_HUMAN  
 ID PHB1\_HUMAN STANDARD; PRT; 199 AA.  
 AC P31941; Q12807;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phorbol 1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Epidermis;  
 RX MEDLINE=99399284; PubMed=10469298;  
 RA Madsen P.P., Anant S., Rasmussen H.H., Gromov P., Vorum H.,  
 RA Dumanski J.P., Tommerup N., Collins J.E., Wright C.L., Dunham I.,  
 RA Macginnitie A.J., Davidson N.O., Cellis J.E.;  
 RT "Psoriasis up-regulated phorbol-1 shares structural but not  
 RT functional similarity to the mRNA-editing protein apobec-1.";  
 RL J. Invest. Dermatol. 113:162-169(1999).  
 RN [2]  
 RP SEQUENCE OF 53-60; 112-121 AND 129-137.  
 RC TISSUE=Keratinocytes;  
 RX MEDLINE=93162043; PubMed=1286667;  
 RA Rasmussen H.H., van Damme J., Fuype M., Gesser B., Cellis J.E.,  
 RA Vandekerckhove J.;  
 RT "Microsequences of 145 proteins recorded in the two-dimensional gel  
 RT protein database of normal human epidermal keratinocytes.";  
 RL Electrophoresis 13:960-969(1992).  
 CC -1- SIMILARITY. BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES  
 CC FAMILY. STRONG, TO APOLIPOPROTEIN B MRNA EDITING PROTEIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; U03891; AAA03706.2;  
 DR Aarhus/Ghent-2DPAGE; 2116; IEF.  
 DR InterPro; IPR002125; dCMP\_cyt\_deam.  
 DR PROSITE; PS00903; CYT\_DCMP\_DEAMINASES; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 199 AA; 23012 MW; 42E99E0D7DF7AA14 CRC64;

Alignment Scores:  
 Pred. No.: 1.87e-24 Length: 199  
 Score: 363.50 Matches: 83  
 Percent Similarity: 58.12% Conservative: 28

Best Local Similarity: 43.46% Mismatches: 57  
 Query Match: 7.35% Indels: 23  
 DB: 1 Gaps: 7

US-09-966-880a-7 (1-2818) x PHB1\_HUMAN (1-199)

QY 92 TTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCTGGCTAAGGTCGG 151  
 DB 12 LeuMetAspProHisIlePheThrSerAsnPheAsnAsn-----GlyIleGlyArg 28  
 QY 152 COTGAGACCTACTGCTGTGCTAGTAGTAAGAGCGGTGACAGTGTACATCTCTTCTC 211  
 DB 29 HisLysThrTyrLeuCysTyrGluValGluArgLeuAspAsnGlyThrSerValLysMet 48  
 QY 212 GAC-----TTTGGTATCTTCGCRAT-----AGAAC-----GGC 241  
 DB 49 AspGlnHisArgGlyPheLeuHisAsnGlnAlaLysAsnLeuLeuCysGlyPheTyrGly 68  
 QY 242 TGCCACGCTGGAATTTCTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGGCCGC 301  
 DB 69 ArgHisAlaGluLeuArgPheLeuAspLeuValProSerLeuGlnLeuAspProAlaGln 88  
 QY 302 TGCTACCGCGCTCACCTGGTTCACCTGCTGAGCCCTGTACGAC-----TGTGCCCGA 355  
 DB 89 IleTyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGly 108  
 QY 356 CATGTGGCGGACTTTCTGCGAGGAGAACCCACCTCAGTCTGAGGATCTTCACCGCGCG 415  
 DB 109 GluValArgAlaPheLeuGlnGluAsnThrHisValArgLeuArgilePheAlaLaArg 128  
 QY 416 CTCCTACTTCTGTGAGGACCGCAGGCTGAGCCC-----GAGGGGCTGGCGGCGCTG 466  
 DB 129 IleTyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeu 143  
 QY 467 CACCGCGCGGGTGCAATGCCATCATGACCTCAAGATATTTTACTGCTGGAAT 526  
 DB 144 ArgAspAlaGlyValGlnValSerIleMetThrTyrAspGluPheLysHisCysTrpasp 163  
 QY 527 ACTTTGTGAGAAACCATCAAGAACTTTCAAGCGCTGGGAGGCGTGCATGAAATTC 586  
 DB 164 ThrPheValAspHisGlnGlyCysProPheGlnProTyrAspGlyLeuAspGluHisSer 183  
 QY 587 GTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 619  
 DB 184 GlnAlaLeuSerGlyArgLeuArgAlaIleLeu 194

RESULT 14  
 ALU3\_HUMAN  
 ID ALU3\_HUMAN STANDARD; PRT; 587 AA.  
 AC P39190;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Alu subfamily SBI sequence contamination warning entry.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M., Makalowski W.;  
 RT "Alu alert";  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie J.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 RT potentially misleading protein sequences.";  
 RL Genomics 12:838-841(1992).  
 RN [3]  
 RP ALU FAMILIES CLASSIFICATION.











GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 14, 2003, 18:12:03 ; Search time 172.791 Seconds  
(without alignments)  
6720.711 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 4948  
Sequence: 1 agagaaccatcataattga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_n2p.model -DEV=rlp  
-Q=/cgn2\_1/USPTO\_spool/US09966880/runat\_14062003\_175524\_10316/app\_query.fasta\_1.9493  
-DB=SPTREMBL\_21 -FMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09966880 -RCGN\_1\_1\_978 @runat\_14062003\_175524\_10316 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	21.9	198	4 Q9GZX7	Q9gzx7 homo sapien

	2	1008	20.4	198	11	Q9WVE0
c	3	403.5	8.2	171	4	Q9H728
	4	390	7.9	384	4	Q9HC16
	5	389	7.9	136	4	Q96NR6
c	6	381.5	7.7	239	4	Q9NX17
c	7	370	7.5	118	4	Q9H387
c	8	369.5	7.5	204	4	Q9Y555
c	9	369	7.5	122	6	Q9BGW3
	10	367.5	7.4	190	4	Q96F12
c	11	360	7.3	133	4	Q96JRS
c	12	357	7.2	127	4	Q9H743
c	13	356.5	7.2	152	4	Q9NX85
c	14	354.5	7.2	294	4	Q9NRW3
	15	350	7.1	429	11	Q99J72
	16	341.5	6.9	386	4	Q96AK3
c	17	339	6.9	115	6	Q9N083
c	18	336	6.8	123	4	Q9HAD8
c	19	321.5	6.5	238	4	Q8WT23
c	20	320.5	6.5	666	4	P78525
c	21	319	6.5	130	4	Q9HBS7
c	22	317	6.4	169	4	Q9H397
c	23	315.5	6.4	139	4	Q9Y4V1
c	24	303	6.1	162	4	Q9BYA5
c	25	303	6.1	165	4	Q96LS9
c	26	303	6.1	375	4	O60448
c	27	298.5	6.0	118	4	Q9P195
c	28	294.5	6.0	368	4	O15662
c	29	292	5.9	214	4	Q9NRE8
c	30	289	5.9	111	4	Q9H3C0
c	31	287	5.8	231	4	Q9NR08
c	32	284	5.7	121	4	Q96N97
c	33	279	5.7	91	4	Q9BYA9
c	34	279	5.7	232	4	Q9H5R3
c	35	278	5.6	224	11	Q9WV35
c	36	277	5.6	90	4	Q96IG1
c	37	275	5.6	535	4	Q96EB1
c	38	274.5	5.6	151	4	Q9HA67
c	39	274	5.5	224	4	Q9Y235
c	40	272	5.5	139	4	Q9BVD9
c	41	272	5.5	163	4	Q96MM0
c	42	270.5	5.5	120	4	Q9H6G8
c	43	270.5	5.5	121	4	Q9Y553
c	44	263	5.3	109	4	Q8WZ39
c	45	261	5.3	83	4	Q96ID7

#### ALIGNMENTS

RESULT 1

Q9GZX7  
ID - Q9GZX7 PRELIMINARY; PRT; 198 AA.  
AC Q9GZX7;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
DE Activation-induced cytidine deaminase.  
GN AID.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20408890; PubMed=10950930;  
RA Muto T., Muramatsu M., Tanikawa K., Kinoshita K., Honjo T.;  
RT "Isolation, tissue distribution and chromosomal localization of the  
RT human activation-induced cytidine deaminase (AID) gene.";  
RL Genomics 68:85-88(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20460541; PubMed=11007475;  
RA Revy P., Muto T., Levy Y., Geissmann F., Plebani A., Sanal O.,  
RA Catalan N., Forvelle M., Dufourcq-Lagelouse R., Gennery A.,

RA Tezcan I., Ersoy F., Kayserili H., Ugazio A.G., Brousse N.,  
RA Muramatsu M., Notarangelo L.D., Kinoshita K., Honjo T., Fischer A.,  
RA Durandy A.;  
RT "Activation-induced cytidine deaminase (AID) deficiency causes the  
RT autosomal recessive form of the Hyper-IgM syndrome (HIGM2).";  
RL Cell 102:565-575(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=B-CELL;  
RA Strausberg R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB040431; BAB12721.1; -;  
DR EMBL; AB040430; BAB12720.1; -;  
DR EMBL; BC006296; AAH06296.1; -;  
DR InterPro; IPR002125; dCMP/cyt\_deam.  
DR PROSITE; PS00903; CYT\_DCMP\_DEAMINASES; UNKNOWN\_1.  
SQ SEQUENCE 198 AA; 23953 MW; 3C27BB143DB184A9 CRC64;

Alignment Scores:  
Pred. No.: 2,65e-105 Length: 198  
Score: 1086.00 Matches: 198  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 21.95% Indels: 0  
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x Q9GZX7 (1-198)

QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCTGG 139  
Db 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
QY 140 GCTAAGGTCGGCGTGAGACCTACCTGTGTAGTGAAGAGGCGTGACGTGTCTACA 199  
Db 21 AlaLysGlyArgargGluThrTyrLeuCysTyrValValLysArgargSsrAlaThr 40  
QY 200 TCCTTTTCTACTGACCTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAAATGCTC 259  
Db 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60  
QY 260 TTCCTCCGCTACATCTCGGACTGGACCTAGACCTGGCGCTGCTACCGCGTCACCTGG 319  
Db 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
QY 320 TTCACCTCTCGGAGCCCTCTGCTACGACTGTGCCCGCATGTGCCGACTTTCTCGGAGGG 379  
Db 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
QY 380 AACCCCACTCAGCTGAGGATCTTCACCGCGCGCTCTACTCTGTGAGGACCGCAAG 439  
Db 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
QY 440 GCTGAGCCGAGGGCTGCGCGGCTGCACCGCGCGGTGCAAAATAGCCATCATGACC 499  
Db 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
QY 500 TTCAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAGAATTTCAAA 559  
Db 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
QY 560 GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGCGATCCTT 619  
Db 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
QY 620 TTGCCCTGTATGAGGTGATGACTTACGAGACGCAATTCGTACTTTGGGACTT 673  
Db 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198

RESULT 2  
Q9WVE0 PRELIMINARY; PRT; 198 AA.  
AC Q9WVE0  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
DE Activation-induced cytidine deaminase.  
GN AICDA OR AID.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99303612; PubMed=10373455;  
RA Muramatsu M., Sankaranand V.S., Anant S., Sugai M., Kinoshita K.,  
RA Davidson N.O., Honjo T.;  
RT "Specific expression of activation-induced cytidine deaminase (AID), a  
RT novel member of the RNA-editing deaminase family in germinal center B  
RT cells.";  
RL J. Biol. Chem. 274:18470-18476(1999).  
DR EMBL; AFI32979; AAD41793.1; -;  
DR MGD; MGI:1342279; Aicda.  
DR InterPro; IPR002125; dCMP/cyt\_deam.  
DR PROSITE; PS00903; CYT\_DCMP\_DEAMINASES; UNKNOWN\_1.  
SQ SEQUENCE 198 AA; 24030 MW; 18A3BA10CA54BEB2 CRC64;

Alignment Scores:  
Pred. No.: 4,51e-97 Length: 198  
Score: 1008.00 Matches: 183  
Percent Similarity: 95.94% Conservative: 6  
Best Local Similarity: 92.89% Mismatches: 8  
Query Match: 20.37% Indels: 0  
DB: 11 Gaps: 0

US-09-966-880A-7 (1-2818) x Q9WVE0 (1-198)

QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCTGG 139  
Db 1 MetAspSerLeuLeuMetLysGlnLysLysPheLeuTyrHisPheLysAsnValArgTrp 20  
QY 140 GCTAAGGTCGGCGTGAGACCTACCTGTGTAGTGAAGAGGCGTGACGTGTCTACA 199  
Db 21 AlaLysGlyArgHisGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
QY 200 TCCTTTTCTACTGACCTTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAAATGCTC 259  
Db 41 SerCysSerLeuAspPheGlyHisLeuArgAsnLysSerGlyCysHisValGluLeu 60  
QY 260 TTCCTCCGCTACATCTCGGACTGGACCTAGACCTGGCGCTGCTACCGCGTCACCTGG 319  
Db 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
QY 320 TTCACCTCTCGGAGCCCTCTGCTACGACTGTGCCCGCATGTGCCGACTTTCTCGGAGGG 379  
Db 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaGluPheLeuArgTrp 100  
QY 380 AACCCCACTCAGCTGAGGATCTTCACCGCGCGCTCTACTCTGTGAGGACCGCAAG 439  
Db 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
QY 440 GCTGAGCCGAGGGCTGCGCGGCTGCACCGCGCGGTGCAAAATAGCCATCATGACC 499  
Db 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleGlyIleMetThr 140  
QY 500 TTCAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAGAATTTCAAA 559  
Db 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnArgGluArgThrPheLys 160  
QY 560 GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGCGATCCTT 619  
Db 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuThrArgGlnLeuArgArgIleLeu 180  
QY 620 TTGCCCTGTATGAGGTGATGACTTACGAGACGCAATTCGTACTTTGGGA 670  
Db 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgMetLeuGly 197



Qy	2189	CCTTGCAGATGCTTCCCA-----TCTCTCTCTCCAAATATGTGTCTCTT	2145
Db	87	ProCysCysSerSerAlaIatpProGluGlySerPheArgProPheGlnMetAsnLeu	106
Qy	2144	TCTCTCTCTCTCTCTTT-----TTTTTTTTTTTTTCTGACACAGAGCTGTGCTCTT	2094

107	PhSerPheLeuSerPhePheLeuPhePhePheLeuArgTrpSerLeuThr-Le	126
2093	GTGCCCGAGCTGGAGTGCACCGCAGGATCTCAGCTTACTGCAACCTCCACCTCCTGGG	2034
126	userProArgLeuGLucysSerAlaIleSerAlaHisCysAsnLeuArgLeuProgl	146
2033	TTCAAGCGATTCTTCCTGCGCTCAGCCTCCCAAGTAGCTGGGATTACAGGTGCTGTGATCCA	1974

Db 146 ySerSerAsnSerProAlaLeuAlaSerGlnValAlaGlyIleThrGlyIleCysHisH1 166

Qy 1973 CGCGTGGCTAAATTTTGGCATATTTGACTACAGACGGGGTTTTGCCATGTTGGCCAGGCTGG 1914  
 166 sAlAr.gInIlePheValIPheLeuValGLuthrGlyPheCysHIsValGLYGLNALAGL 186  
 Db

**QY** 1913 TCTCAACCTCCTCACACAGGTGATCCGCCGGCTCGGCCTCCCAAAGTCTGGGATTAC 1854  
|||||:||||| :||| ||| ||||| ||||| |||||  
**Db** 186 YLeuGIuLeuIleSerGlyAspSerProAlaSerAlaPheGlnSerAlaGlyIle|| 206

Oy		1853	AGGCGTGAGCCACCAACCCTGGGCC---	TCTCCTTTCTT	1817
Dd		206	eGlyValSerHisArgAlaArgProGlySerValPheLeu		219

RESULT 7	
Q9H387	
ID Q9H387	PRELIMINARY;
AC Q9H387;	PRT; 118 AA.

DT	01-MAR-2001	(TREMBLrel. 16, Created)
DT	01-MAR-2001	(TREMBLrel. 16, Last sequence update)
DT	01-MAR-2001	(TREMBLrel. 16, Last annotation update)
DE	PRO2550.	

OC Homo sapiens (human) ;  
 OC Eukaryota; Metazoa;  
 OC Chordata;  
 OC Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;

RC TISSUE=LIVER;  
RA Zhang C., Yu Y., Zhang S., Wei H., Bi J., Zhou G., Dong C., Zai Y.,

<sup>RT</sup> "Functional prediction of the coding sequences of 75 new genes deduced  
<sup>RL</sup> by analysis of cDNA clones from human fetal liver.";  
<sup>RT</sup> Submitted (Feb-1999) to the ENBL/GenBank/DBJ databases.  
<sup>DR</sup> FMRI : AF130080, NC025515.

SQ SEQUENCE 118 AA; 13257 MW; 94688870CAC8760D CRC64;

Alignment Scores:

Pred. No.:	8	39a-10	Length:	110
------------	---	--------	---------	-----

Score:	370.00	Matches:	77
--------	--------	----------	----



```

Db      44  uLysGlnPheSerCysLeuSerLeuProSerSerTrpAsnTyrArgCysLeuProProH1 64
QY      1971  CTTGGCTAAATTTTGGCATGACAGCGGGTTTCCCATGTTGGCCAGCTGGTC 1912
Db      64  sLeuAlaLysPheCysIlePheSer-Argasp-----TyrValGlyGlnAlaGlyL 81
QY      1911  TCAAACTCTCTGACACAGGTGATCCGCCCGCTCGGCTCCCAAGTGTCTGGGATTACAG 1852
Db      81  euGluLeuLeuThrSerGlyAsnProLeuAlaSerAlaSerGlnSerAlaGlyIleThrG 101
QY      1851  GCCTGAGCCACACGCGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1811
Db      101  lyMetSerHisHisSerArgProGlnMetPheLeuGluPhe 114

RESULT 10
Q96F12
ID      Q96F12      PRELIMINARY;      PRT;      190 AA.
AC      Q96F12;
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE      Similar to APOBEC1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=UTERUS;
RA      Strausberg R.;
RL      Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC011739; AAH11739.1; -.
DR      InterPro; IPR002125; dCMP/cytL_deam.
DR      PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ      SEQUENCE 190 AA; 22827 MW; DA0584EF75C91CF0 CRC64;

```

```

Alignment Scores:
Pred. No.:      1,75e-29      Length:      190
Score:          367.50      Matches:      79
Percent Similarity: 58.52%      Conservative: 24
Best Local Similarity: 44.89%      Mismatches: 64
Query Match:      7.43%      Indels: 9
DB:              4      Gaps: 4

```

US-09-966-880A-7 (1-2818) x Q96F12 (1-190)

```

QY      110  TTTCTTTTACCAATTCAAAATGTCCTGGCTAAGGTCGGCGTGAGACCTACCTGTGC 169
Db      17  PheTyrPheGlnPheLysAsnLeuTrpGluAlaAsnArgAspGluThrTrpLeuCys 36
QY      170  TAGCTAGTGAAG-----AGGCGTGACAGTGCTACATCCTTTTCACCTGGACTTGGTTAT 223
Db      37  PheThrValCluGlyIleLysArgArgSerValValSerTrpLysThr-----GlyVal 54
QY      224  CTTCCGAAT-----AGAACGGCTGCCAGTGGAATGTCTTCCTCCGCTAC 271
Db      55  PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74
QY      272  ATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGCTACCTGGCTACCTGGTCCCTGG 331
Db      75  PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTyrSerTrp 94
QY      332  AGCCCTGCTAGCATGTGCCCGGACATGTGGCGGACTTTCTGGGAGGGAACCCCAACCTC 391
Db      95  SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 114
QY      392  AGTCTGAGGATCTTACCGGGCCCTCTACTTCTGTGAGACCGCAGGCTGAGCCCGAG 451
Db      115  AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu 133
QY      452  GGGCTCGGGCGGTGCACCGCGCGGGTGCRAATAGCCATCATGACCTTCAAAGATTAT 511

```

A

```

Db      134  GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluLeuMetAspTyrGluAspPhe 153
QY      512  TTTTACTGCTGGAAATCTTTTGTAGAAAACCATGAAAGAACTTTCAAAGCTCGGAAGGG 571
Db      154  LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly 173
QY      572  CTCGATGAAATACAGTTCTCTCTCCAGACAGCTTCGGCGCATCCTT 619
Db      174  LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu 189

RESULT 11
Q96JR5
ID      Q96JR5      PRELIMINARY;      PRT;      133 AA.
AC      Q96JR5;
DT      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE      PRO0764.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=LIVER;
RX      MEDLINE=21376145; PubMed=11483580;
RA      Yu Y., Zhang C., Zhou G., Wu S., Ou X., Wei H., Xing G., Dong C.,
RA      Zhai Y., Wan J., Ouyang S., Li L., Zhang S., Zhou K., Zhang Y., Wu C.,
RA      He F.;
RT      "Gene Expression Profiling in Human Fetal Liver and Identification of
RT      Tissue- and Developmental-Stage-Specific Genes through Compiled
RT      Expression Profiles and Efficient Cloning of Full-Length cDNAs.";
RL      Genome Res. 11:1392-1403(2001).
DR      EMBL; AF305818; AAK55521.1; -.
SQ      SEQUENCE 133 AA; 14414 MW; C9318E532FB139F CRC64;

```

Alignment Scores:

```

Pred. No.:      9.83e-29      Length:      133
Score:          360.00      Matches:      68
Percent Similarity: 82.98%      Conservative: 10
Best Local Similarity: 72.34%      Mismatches: 16
Query Match:      7.29%      Indels: 0
DB:              4      Gaps: 0

```

US-09-966-880A-7 (1-2818) x Q96JR5 (1-133)

```

QY      2071  GGCACATCTACGCTTACTCCAACTCCACCTCTCGGTTCAGCGCATTCCTCGCTCA 2012
Db      29  GlySerLeuGlnThrHisCysAsnIleHisLeuProGlySerSerAspProAlaSer 48
QY      2011  GCCTCCCAAGTAGCTGGGATTACAGCTGCTGTACAGCGCTGGCTAATTTTGCATTT 1952
Db      49  AlaSerTrpValAlaGlyIleThrGlyThrArgHisHisThrTrpLeuIlePheValPhe 68
QY      1951  TGAGTACAGACGGGGTTTGGCCATGTTGGCAGGCTGGTCTCAAACTCTGACACAGGT 1892
Db      69  LeuValValThrGlyPheHisHisValGlyGlnAlaGlyLeuLysLeuLeuThrSerGly 88
QY      1891  GATCCCGCGCTCGGCTCCCAACTGCTGGATTACAGCGGTAGCCAGCCACCGCCCG 1832
Db      89  AspProProSerSerSerGlnSerAlaGlyIleThrGlyValSerHisHisAlaGln 108
QY      1831  CCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1790
Db      109  ProTyrSerCysIlePhePhePhePhePhePheGluMetGlu 122

```

RESULT 12

```

Q9H743
ID      Q9H743      PRELIMINARY;      PRT;      127 AA.
AC      Q9H743;
DT      01-MAR-2001 (TrEMBLrel. 16, Created)
DT      01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT      01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

```



[illegible]

```

QY 110 TTTCTTACCAATTCAAAATGTCGCTGGCTAGGGTCGCGTGAGACCTACCTGTGC 169
Db 17 PheTyrPheGlnPheLeuSerGluAlaAsnAspArgAsnGluThrTrpLeuCys 36
QY 170 TACGTAGTGAAG-----AGGCGTGACAGTGCTACATCTTTTCTACTGGACTTTGGTTAT 223
Db 37 PheThrValGluGluGlyLeuSerValValSerTrpLysThr-----GlyVal 54
QY 224 CTTCGCAAT-----AAGAACGGCTGCCACGTGGGAATGCTCTCTCTCCGCTAC 271
Db 55 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74
QY 272 ATCTCGGACTGGGACCTAGACCCGCGCTGCTACCGCGTGACCTGGTTCACCTCTGG 331
Db 75 PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTyrThrSerTrp 94
QY 332 AGCCCTCTACGACTGTGCCCGACATGTGCGGACCTTTCTCGGAGGAAACCCCAACCTC 391
Db 95 SerProCysProAspCysAlaGlyValAlaGluPheLeuAlaArgHisSerAsnVal 114
QY 392 AGTCTGAGGATCTTCACGCGCGCTCTACTTCTGTGAGGACCGACGCTGAGCCCGAG 451
Db 115 AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu 133
QY 452 GGGCTGGCGGCTGCACCGCGCGGCTGCAATAGCCATCATGCTTCAAGATTAT 511
Db 134 GlyLeuArgSerLeuSerGlnGluGlyValAlaGluIleMetAspTyrGluAspPhe 153
QY 512 TTTTACTGCTGGAATCTTTGTAGAAACCATGAAAGACTTTTCAAGCTGGGAAGGG 571
Db 154 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpGluGly 173
QY 572 CTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCTTTTGGCCCTGTAT 631
Db 174 Ile---LysAsnGlnLeuSerThrSerGluLysLysAlaThrGlyGluSerProValArg 192
QY 632 GAGGTTGATGACTTACGAGAGCGCATTTTCGTTACTTTGGGACTTTGATACCACTTCCAGGA 691
Db 193 -----GlyLeuProGly 196
QY 692 ATGTCACACAGATGAATATCTCTGCTGAGACAGTGGATAAAAAACAGTCTTCAAGT 751
Db 197 ---ProHisGlyLeuSerProLeuAlaSerCysSerCysThrGlyLeuProSerThr 215
QY 752 CTT-----CTCTGTTT-----TATPCT 769
Db 216 LeuAspProLeuCysPheCysLeuValIleLeuSerProSerTrpProGlnGlyHisSer 235
QY 770 TCAACTCTCACTTTCTTAGAGTTTACAGAAAAAATATT 808
Db 236 ThrValLeuProCysLeuThrAlaSerSerLeuPhe 248

```

## RESULT 15

```

Q99J72 ID Q99J72 PRELIMINARY; PRT; 429 AA.
AC Q99J72;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 51.0 kDa protein.
GN BC003314.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003314; AA003314.1;
DR MGD; MGI:1933111; BC003314.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_2.

```

KW Hypothetical protein.  
SQ SEQUENCE 429 AA; 51017 MW; BE44D01380AD7F6E CRC64;

## Alignment Scores:

Pred. No.: 1.52e-27 Length: 429  
Score: 350.00 Matches: 81  
Percent Similarity: 49.39% Conservative: 40  
Best Local Similarity: 33.06% Mismatches: 84  
Query Match: 7.07% Indels: 40  
DB: 11 Gaps: 6

US-09-966-880A-7 (1-2818) x Q99J72 (1-429)

```

QY 92 TTGTATGACCCGGAGGAGAGTCTTTTACCAATTCAAAATGTCGCGTGGCTAAGGTCGG 151
Db 20 LeuIleSerGlnGluThrPheLysPheHisPheLysAsnLeuArgTyrAlaIleAspArg 39
QY 152 CGTGAGACTACCTGTCTAGTAGTGAAGAGGCGTACAGTGTACATCTCTTTTCTACTG 211
Db 40 LysAspThrPheLeuCysTyrGluValThrArgLysAspCysAspSerProValSerLeu 59
QY 212 GACTTTGGTTATCTTCGCAATAGACGGCTGCCACGTGGAAATGCTTCTCTCCGCTAC 271
Db 60 HisHisGlyValPheLysAsnLysAspAsnIleHisAlaGluIleCysPheLeuTyrTrp 79
QY 272 ATCTCGGAC-----TGGGACCTAGACCTGGCGCTGCTACCGCGCTCACCTGGTTC 322
Db 80 PheHisLysValLeuLysValLeuSerProArgGluGluPheLysIleThrTrpTrp 99
QY 323 ACCTCTGGAGCCCTGCTACGACTGTGCCCGACATGTGCCGAGTCTTCTCGGAGGCAAC 382
Db 100 MetSerTrpSerProCysPheGluCysAlaGluValLeuArgPheLeuAlaThrHis 119
QY 383 CCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGAGCCGCAAGCT 442
Db 120 HisAsnLeuSerLeuAspIlePheSerArgLeuTyrAsnIleArgAspProGluAsn 139
QY 443 GAGCCCGAGGGGTGCGCGGCTGCACCGCGCGGTGCAAAATAGCATCATGACCTTC 502
Db 140 Gln---GlnAsnLeuCysArgLeuValGlnGluGlyAlaGlnValAlaAlaMetAspLeu 158
QY 503 AAGATTATTTTACTGCTGGAATCTTTTGTAGAAACCATGAAAGAACTTTTCAAGGCC 562
Db 159 TyrGluPheLysLysCysTrpLysLysPheValAspAsnGlyGlyArgPheArgPro 178
QY 563 TGGGAAGGGGTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCCCATCTTTTG 622
Db 179 TrpLysLysLeuLeuThrAsnPheArgTyrGlnAspSerLysLeuGlnGluIleLeuArg 198
QY 623 CCCTCTGATGAGGTTGATGACTTTACGAGACGCACTTTCGTACTTTGGGACTTTTGATGCA 682
Db 199 ProCysTyr----- 201
QY 683 CTTCGA-----GGAATGTCACACAGATGAATATCTCTGTGGAACACAGTGGATA 733
Db 202 IleProValProSerSerSerSerSerThrLeuSerAsnIleCys-----LeuThr 218
QY 734 AAAACAGCTCTTCAAGTCTTCTCTGT----- 760
Db 219 LysGlyLeuProGluThrArgPheCysValGluGlyArgArgValHisLeuLeuSerGlu 238
QY 761 -----TTTTATTCT 769
Db 239 GluGluPheTyrSer 243

```

Search completed: June 14, 2003, 18:47:39  
Job time : 181.791 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 14, 2003, 18:13:03 ; Search time 27.1417 Seconds  
(without alignments)  
6109.704 Million cell updates/sec

Title: US-09-966-880A-7  
Perfect score: 4948  
Sequence: 1 agagaccacatcataattga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US0966880/runat\_14062003\_175525\_10370/app\_query.fasta\_1.9493  
-DB=issued\_Patents\_AA -QfMT=fastan -SUFFIX=n2p.ra -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cd1  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USPR=US0966880.cgnl\_1.139.runat\_14062003\_175525\_10370 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/1/iaa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/iaa/5B.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/iaa/6A.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/iaa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	369.5	7.5	190	1	US-08-816-241-1
2	369.5	7.5	190	3	US-09-128-395-1
C 3	332	6.7	132	4	US-09-605-785-573
C 4	303	6.1	375	2	US-08-454-557C-121
C 5	303	6.1	375	2	US-08-340-426D-121
C 6	303	6.1	375	2	US-08-450-673C-121
C 7	295	6.0	397	5	PCT-US95-17111A-121
C 8	287	5.8	1079	3	US-09-058-489-22
C 9	286	5.8	99	4	US-09-288-143-168
C 10	278	5.6	110	4	US-09-227-357-193
11	271	5.5	222	1	US-08-687-895-1
12	271	5.5	222	2	US-09-040-482-1

C 13	248	5.0	122	4	US-09-227-357-285	Sequence 285, App
14	245	5.0	116	1	US-08-687-895-3	Sequence 3, Appli
15	245	5.0	116	1	US-08-816-241-3	Sequence 3, Appli
16	245	5.0	116	2	US-09-040-482-3	Sequence 3, Appli
17	245	5.0	116	3	US-09-128-395-3	Sequence 3, Appli
C 18	225.5	4.6	90	4	US-09-227-357-171	Sequence 171, App
19	221	4.5	236	1	US-08-687-895-4	Sequence 4, Appli
20	221	4.5	236	1	US-08-816-241-4	Sequence 4, Appli
21	221	4.5	236	2	US-09-040-482-4	Sequence 4, Appli
22	221	4.5	236	3	US-09-128-395-4	Sequence 4, Appli
C 23	218.5	4.4	56	4	US-09-227-357-577	Sequence 577, App
24	214.5	4.3	236	1	US-08-158-682A-4	Sequence 4, Appli
25	199.5	4.0	229	1	US-08-158-682A-2	Sequence 2, Appli
26	199.5	4.0	229	1	US-08-015-203-2	Sequence 2, Appli
27	199.5	4.0	229	1	US-08-687-895-5	Sequence 5, Appli
28	199.5	4.0	229	1	US-08-816-241-5	Sequence 5, Appli
29	199.5	4.0	229	2	US-09-040-482-5	Sequence 5, Appli
30	199.5	4.0	229	3	US-09-128-395-5	Sequence 5, Appli
31	191.5	3.9	76	4	US-09-605-785-575	Sequence 575, App
C 32	165.5	3.4	500	4	US-09-265-630-13	Sequence 13, Appl
C 33	160	3.2	72	4	US-09-227-357-655	Sequence 655, App
C 34	157	3.2	368	4	US-09-265-630-11	Sequence 11, Appl
C 35	156.5	3.2	579	3	US-08-704-711A-1	Sequence 1, Appli
C 36	145.5	3.2	579	4	US-09-521-220-1	Sequence 1, Appli
37	145.5	2.9	52	4	US-09-227-357-537	Sequence 537, App
38	134	2.7	447	1	US-08-365-689-1	Sequence 1, Appli
39	134	2.7	447	1	US-07-747-781-1	Sequence 1, Appli
40	134	2.7	447	1	US-08-145-138A-1	Sequence 1, Appli
41	134	2.7	447	5	PCT-US92-08688-1	Sequence 1, Appli
C 43	128	2.6	47	4	PCT-US93-03027-5	Sequence 5, Appli
C 44	120	2.4	48	3	US-09-227-357-656	Sequence 656, App
C 45	120	2.4	619	3	US-08-951-200A-1	Sequence 1, Appli
					US-09-156-253-48	Sequence 48, Appl

#### ALIGNMENTS

RESULT 1  
US-08-816-241-1  
; Sequence 1, Application US/08816241  
; Patent No. 5804185  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/816,241  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0239 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 1:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT09
; CLONE: 1646823
US-08-816-241-1

Alignment Scores:
Pred. No.: 2,94e-32 Length: 190
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 7.47% Indels: 9
DB: 1 Gaps: 4

US-09-966-880A-7 (1-2818) x US-08-816-241-1 (1-190)
QY 110 TTTCCTTTACCAATTCAAAATGTCGCTGGCTAAGGTCGGCTGAGACCTACCTGTGC 169
Db 17 PheTyrPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys 36
QY 170 TAGCTAGTGAAG-----AGGCGTCAGAGTGTACATCCCTTTTCACTGGACTTTGGTTAT 223
Db 37 PheThrValGluGlyIleLysArgArgSerValValSerTrpLysThr-----GlyVal 54
QY 224 CTTCGCAAT-----AAGAACGGCTGCCACGTGGGAATTCCTTCTCCGCTAC 271
Db 55 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74
QY 272 ATCTCGGACTGGGACCTAGACCTGACCTGCGCTGCTACCGCTACCTGTTTCACTGCTGG 331
Db 75 PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTyrThrSerTrp 94
QY 332 AGCCCTCTAGTACGACTGTCGCGGACATGTGCGCCGCTTTCGCGAGGAGCAACCACTC 391
Db 95 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 114
QY 392 AGTCTGAGGATCTTTTGTAGAAACCATGCAATGCCATCATGACCTTCAAGATTAT 511
Db 115 AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu 133
QY 452 TTTTACTGCTGGAATCTTTGTAGAAACCATGCAATGCCATCATGACCTTCAAGATTAT 511
Db 134 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrGluAspPhe 153
QY 512 TTTTACTGCTGGAATCTTTTGTAGAAACCATGCAATGCCATCATGACCTTCAAGATTAT 511
Db 154 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly 173
QY 572 CTGCATGAAATTCAGTTCCTCTCCACAGACTTCGCGCATCCTT 619
Db 174 LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu 189

RESULT 2
US-09-128-395-1
; Sequence 1, Application US/09128395
; Patent No. 6087108
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,395
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/816,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT09
; CLONE: 1646823
US-09-128-395-1

Alignment Scores:
Pred. No.: 2,94e-32 Length: 190
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 7.47% Indels: 9
DB: 1 Gaps: 4

US-09-966-880A-7 (1-2818) x US-09-128-395-1 (1-190)
QY 110 TTTCCTTTACCAATTCAAAATGTCGCTGGCTAAGGTCGGCTGAGACCTACCTGTGC 169
Db 17 PheTyrPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys 36
QY 170 TAGCTAGTGAAG-----AGGCGTCAGAGTGTACATCCCTTTTCACTGGACTTTGGTTAT 223
Db 37 PheThrValGluGlyIleLysArgArgSerValValSerTrpLysThr-----GlyVal 54
QY 224 CTTCGCAAT-----AAGAACGGCTGCCACGTGGGAATTCCTTCTCCGCTAC 271
Db 55 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74
QY 272 ATCTCGGACTGGGACCTAGACCTGACCTGCGCTGCTACCGCTACCTGTTTCACTGCTGG 331
Db 75 PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTyrThrSerTrp 94
QY 332 AGCCCTCTAGTACGACTGTCGCGGACATGTGCGCCGCTTTCGCGAGGAGCAACCACTC 391
Db 95 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 114
QY 392 AGTCTGAGGATCTTTTGTAGAAACCATGCAATGCCATCATGACCTTCAAGATTAT 511
Db 115 AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu 133
QY 452 TTTTACTGCTGGAATCTTTGTAGAAACCATGCAATGCCATCATGACCTTCAAGATTAT 511
Db 134 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrGluAspPhe 153
QY 512 TTTTACTGCTGGAATCTTTTGTAGAAACCATGCAATGCCATCATGACCTTCAAGATTAT 511
Db 154 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly 173
QY 572 CTGCATGAAATTCAGTTCCTCTCCACAGACTTCGCGCATCCTT 619

```

Db 174 LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu 189

RESULT 3

US-09-605-785-573

; Sequence 573, Application US/09605785

; Patent No. 6321716

; GENERAL INFORMATION:

; APPLICANT: Xu, Jiangchun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Jiang, Yuqul

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.

; APPLICANT: Day, Craig H.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Carter, Darrick

; APPLICANT: Li, Samuel

; APPLICANT: Wang, Aijun

; APPLICANT: Skelky, Yasir A.W.

; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.427C16

; CURRENT APPLICATION NUMBER: US/09/605,785

; CURRENT FILING DATE: 2000-06-27

; NUMBER OF SEQ ID NOS: 835

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 573

; LENGTH: 132

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-605-785-573

Alignment Scores:

Pred. No.: 3 51e-28 Length: 132

Score: 332.00 Matches: 65

Percent Similarity: 71.43% Conservative: 5

Best Local Similarity: 66.33% Mismatches: 28

Query Match: 6.73% Indels: 0

DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-605-785-573 (1-132)

QY 2123 TTTTCTTCTGAGACAGAGTCTTGCTGTGCGCCAGCGTGGAGTGCACGCGCAGAT 2064

Db 28 PhePhePheLeuArgGlnGluSerGlyProValAlaGlnAlaGlyValGlnTrpHisAsp 47

QY 2063 CTCAGCTTACTGCAACCTCCACCTCTCTGGTTCACGCGATTCTCTGCTCAGCTCCCA 2004

Db 48 LeuSerSerLeuGlnProLeuProHisArgPheLeuGlnPheSerCysLeuSerLeuPro 67

QY 2003 AGTAGCTGGATTACAGGTGCGCTGTACACAGCGCTGGCTAAATTTTTCATTTTGACTACA 1944

Db 68 HisSerTrpAspHisArgTrpAlaProProHisLeuAlaAsnPheCysSerPheSerArg 87

QY 1943 GACGGGGTTTGGCCATGTTGGCCAGGCTGGTCTCAAACTCCTGACACAGGTGATCCGCC 1884

Db 88 AspGlyValSerLeuLysCysSerGlyTrpSerLysThrProGlyLeuGlnSerAla 107

QY 1883 CGGCTCGGCTCCCAAGTGTGGGATTACAGCGTGGACCGCAGCCAGCCCGGCC 1830

Db 108 CysLeuGlyLeuProLysCysTrpGlyTyrArgHisLysProProHisProAla 125

RESULT 4

US-08-454-557C-121

; Sequence 121, Application US/08454557C

; Patent No. 5830670

; GENERAL INFORMATION:

; APPLICANT: de la Monte, Suzanne

; APPLICANT: Wands, Jack R.

; TITLE OF INVENTION: Neural Thread Protein Gene Expression and Detection

; NUMBER OF SEQUENCES: 121

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

; STREET: 1100 New York Avenue, Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; APPLICATION NUMBER: US/08/454,557C

; FILING DATE: 30-MAY-1995

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Ludwig, Steven R.

; REGISTRATION NUMBER: 36,203

; REFERENCE/DOCKET NUMBER: 0609.3840003

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202) 371-2600

; TELEFAX: (202) 371-2540

; INFORMATION FOR SEQ ID NO: 121:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 375 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-454-557C-121

Alignment Scores:

Pred. No.: 1 05e-24 Length: 375

Score: 303.00 Matches: 90

Percent Similarity: 47.14% Conservative: 9

Best Local Similarity: 42.86% Mismatches: 41

Query Match: 6.14% Indels: 71

DB: 2 Gaps: 6

US-09-966-880A-7 (1-2818) x US-08-454-557C-121 (1-375)

QY 2098 CTCCTTGTCCCGCAGGCTGGAGTGCACAGCGCAGCATCTCAGCTTACTGCAACCTCCACCTC 2039

Db 5 LeuLeuLeuProArgLeuGluCysAsnGlyAlaIleSerAlaHisArgAsnLeuArgLeu 24

QY 2038 CTGGGTTCAAGCGATTCTCTGCTCAGCTCCCAAGTAGCTGGGATTACAGGTGCGCTGC 1979

Db 25 ProGlySerSerAspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCys 44

QY 1978 TACCACGCTGGCTAAAT---TTTGCATTTTGTAGTACAGACGGGGTTTGGCCATGTGGC 1922

Db 45 ThrHisAlaArgLeuIleLeuTyrPhePheLeuValGluMetGluPheLeuHisValGly 64

QY 1921 CAGGCTGGCTCAAACTCCTGACACAGGTGATCCGCCGG----- 1881

Db 65 GlnAlaGlyLeuGluLeuProThrSerAspProSerValSerAlaSerGlnSerAla 84

QY 1881 ----- 1881

Db 85 ArgTyrArgThrGlyHisHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArg 104

QY 1880 -----CTCGGCCCTC 1872

Db 105 ValSerLeuMetCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeu 124

QY 1871 CCAAAGTGTGGATTACAGGCGTGAGCCACGACCGCCCGCTCTCTCTTTCTTTT 1812

Db 125 ProLysCysTrpAspTyrArgArg--AlaAlaValProGly-----LeupheI 140

[illegible]

; LENGTH: 375 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-450-673C-121

Alignment Scores:  
Pred. No.: 1.05e-24 Length: 375  
Score: 303.00 Matches: 90  
Percent Similarity: 47.14% Conservative: 9  
Best Local Similarity: 42.88% Mismatches: 41  
Query Match: 6.14% Indels: 71  
DB: 2 Gaps: 6

US-09-966-880A-7 (1-2818) x US-08-450-673C-121 (1-375)

QY 2098 CTCCTGTCGCCAGCGCTGGAGTGGAAAGCGCAGCATCTAGCTTACTGCAACCTCCACCTC 2039  
Db 5 LeuLeuLeuProArgLeuGluCysAsnGlyAlaIleSerAlaHisArgAsnLeuArgLeu 24  
QY 2038 CTGGGTTCAAGCGATTCTCCTGCCCTCAGCCTCCCAAGTAGCTGGGATTACAGCTGCCCTGC 1979  
Db 25 ProGlySerSerAspSerProAlaSerAlaSerProValAlaGlyIleThrGlyMetCys 44  
QY 1978 TACCACGCTGGCTAATT---TTTGCATTTTGGTACAGACGGGGTTTGGCCATGTTGCC 1922  
Db 45 ThrHisAlaArgLeuIleLeuTyrPhePheLeuValGluMetGluPheLeuHisValGly 64  
QY 1921 CAGGCTGTCTCAAACTCTGACACAGGTGATCGCGCGG----- 1881  
Db 65 GlnAlaGlyLeuGluLeuProThrSerAspAspProSerValSerAlaSerGlnSerAla 84  
QY 1881 ----- 1881  
Db 85 ArgTyrArgThrGlyHisAlaArgLeuCysLeuAlaAsnPheCysGlyArgAsnArg 104  
QY 1880 -----CTCGGCTC 1872  
Db 105 ValSerLeuMetCysProSerTrpSerProGluLeuLysGlnSerThrCysLeuSerLeu 124  
QY 1871 CCAAGTCTGGGATTACAGGCTGAGCCAGCACCACCGCGCCTCTCTCTTTCTTTT 1812  
Db 125 ProLysCysTrpAspTyrArgArg--AlaAlaValProGly-----LeuPheI 140  
QY 1811 TTTTCTTTTGTGAGACAGCATCTGTTATGTTGCCA----- 1771  
Db 140 LeuPhePheLeuArgHisArg-----CysProThrLeuThrGlnAspGluV 156  
QY 1770 --GGCTGG-----CCTGAACACCTGTGTTCAAGCCATCCTCC 1734  
Db 156 alGlnTrpCysAspHisSerSerLeuGlnProSerThrProGluIleLys-HisProPro 175  
QY 1733 TCCTCAACTCCCAAGCAGCTAGACTATGACGCTGTCGCCAGGTTGCCAGCTTATCAT 1674  
Db 176 AlaSerAlaSerGlnValAlaGlyThrLysAspMetHisHisTyrThrTriLeu-IlePh 195  
QY 1673 TCTGTTTGTCACTCTCTT 1652  
Db 195 eilePheIlePheAsnPheLeu 202

RESULT 7

PCT-US95-17111A-121  
; Sequence 121, Application PC/TUS9517111A  
; GENERAL INFORMATION:  
; APPLICANT: de la Monte, Suzanne  
; APPLICANT: Wands, Jack R.  
; TITLE OF INVENTION: Neural Thread Protein Gene Expression and  
; TITLE OF INVENTION: Detection of Alzheimer's Disease  
; NUMBER OF SEQUENCES: 121  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.  
; STREET: 1100 New York Avenue, Suite 600  
; CITY: Washington

; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/17111A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/340,426  
; FILING DATE: 14-NOV-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Ludwig, Steven R.  
; REGISTRATION NUMBER: 36,203  
; REFERENCE/DOCKET NUMBER: 0609.3840002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 121:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 397 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US95-17111A-121

Alignment Scores:  
Pred. No.: 8.4e-24 Length: 397  
Score: 295.00 Matches: 69  
Percent Similarity: 66.67% Conservative: 11  
Best Local Similarity: 57.50% Mismatches: 34  
Query Match: 5.98% Indels: 6  
DB: 5 Gaps: 1

US-09-966-880A-7 (1-2818) x PCT-US95-17111A-121 (1-397)

QY 2181 TCGTTCCTC-----ATCCTTCTCTCCAAATATGTCTCTTCTCTCTCTCTCT 2134  
Db 277 CysLeuProArgProProLysValLeuGlyLeuGlnAspValThrProThrAlaArgPro 296  
QY 2133 CTTCTTTTCTTTTCTTTTCTCGACAGAGTCTGCTCTGTCGCCAGCTCGAGTGCA 2074  
Db 297 IlePheAsnPheCysLeuPheGluMetGluSerHisSer-ValThrGlnAlaGlyValGI 316  
QY 2073 ACGGCAGCATCTCAGCTTACTGCACCTCCACCTCCTGGGTTCACAGCGATTCTCTCCCT 2014  
Db 316 nTrpProAsnLeuGlySerLeuGlnProLeuProGlyLeuLysArgPheSerCysLe 336  
QY 2013 CAGCCTCCCAAGTAGCTGGGATTACAGTGCCTGCTTACACAGCTCGCTTAATTTTGCAT 1954  
Db 336 uSerLeuProSerSerTrpAspTyr-GlyHisLeuHisHisThrProLeuIlePheValP 356  
QY 1953 TTTGAGTACAGACGGGTTTTCCTGCTGGCAGCTGCTCTCAACCTCCACACAG 1894  
Db 356 heSerLeuGluAlaGlyPheHisHisIleCysGlnAlaGlyLeuLysLeuLeuThrSerg 376  
QY 1893 GTGATCCGCGCTCGGCTCCCAAGTGCTGGGATTACAGCGCTGAGCCACAC 1838  
Db 376 lyAspProProAlaSerAlaPheGlnSerAlaGlyIleThrGlyValThrProHis 394

RESULT 8

US-09-058-489-22  
; Sequence 22, Application US/09058489  
; Patent No. 6103886  
; GENERAL INFORMATION:  
; APPLICANT: Whitehead Institute for Biomedical Research  
; APPLICANT: Lahn, Bruce  
; APPLICANT: Page, David  
; TITLE OF INVENTION: Genes in the No. 6103886-Recombining Region of

```

; TITLE OF INVENTION: the Y Chromosome
; FILE REFERENCE: WHI97-089A
; CURRENT APPLICATION NUMBER: US/09/058,489
; CURRENT FILING DATE: 1998-04-10
; EARLIER APPLICATION NUMBER: 60/041,877
; EARLIER FILING DATE: 1997-04-11
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 22
; LENGTH: 1079
; TYPE: prnt
; ORGANISM: Human
US-09-058-489-22

```

Alignment Scores:		
Pred. No.:	1.14e-22	Length:
Score:	287.00	Matches:
Percent Similarity:	61.54%	Conservative:
Best Local Similarity:	53.15%	Mismatches:
Query Match:	5.81%	Indels:
DB:	3	Gaps:
		5
		1079

US-09-966-880A-7 (1-2818) x US-09-058-489-22 (1-1079)

[illegible]

```

RESULT 9
US-09-288-143-168
; Sequence 168, Application US/09288143
; Patent No. 6433139
; GENERAL INFORMATION:
; APPLICANT: Brewer et al.
; TITLE OF INVENTION: 53 Human Secreted Proteins
; FILE REFERENCE: PZ018P1
; CURRENT APPLICATION NUMBER: US/09/288,143
; CURRENT FILING DATE: 1999-04-08
; EARLIER APPLICATION NUMBER: PCT/US98/21142
; EARLIER FILING DATE: 1998-10-08
; EARLIER APPLICATION NUMBER: 60/061,453
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/061,529
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: 60/071,498

```

```

? EARLIER FILING DATE: 1997-10-09
? EARLIER APPLICATION NUMBER: 60/061,527
? EARLIER FILING DATE: 1997-10-09
? EARLIER APPLICATION NUMBER: 60/061,536
? EARLIER FILING DATE: 1997-10-09
? EARLIER APPLICATION NUMBER: 60/061,532
? EARLIER FILING DATE: 1997-10-09
? NUMBER OF SEQ ID NOS: 219
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 168
? LENGTH: 99
? TYPE: PRT
? ORGANISM: Homo sapiens
? US-09-288-143-168

```

Alignment Scores:	
Pred. No.:	3.84e-23
Score:	286.00
Length:	99
Matches:	63
Percent Similarity:	71.88%
Conservative:	27
Best Local Similarity:	65.62%
Mismatches:	27
Query Match:	5.79%
Indels:	1
DB:	4
Gaps:	0
DB:	0

US-09-966-880A-7 (1-2818) x US-09-288-143-168 (1-99)

	Qy	2094	TGTGCGCCAGCGTGGAGTGCAACGCCAGCATCTCAGCTTACTGCACCTCCACCTCTCTGG	2035
	Db	2	CysCysProGlyTrpSerAlaValAArgSerTrpLeuThrAlaThrLeuAlaSerTrp	21
	Qy	2034	GTTCAACGGATTCTTCCTGCC - TCAGCCTCCCAGTAGCTGGGATTCACAGTTCGGCTGCATC	1976
	Db	22	ValGlnAlaIleuMetaspSerAlaSerGlnValAlaglyllethrServalHISHis	41
	Qy	1975	CACGCCGTGGCTAAATTGTGATTGTAGTACAGACGGGGTTTTGCCATGTGGCCAGGCT	1916
	Db	42	GlnAlaGlnLeuSerPheValPheLeuValGlumetGlyLeuCysHisValGlyGlnAla	61
	Qy	1915	GGTCTCAAAGTCTGCACCACAGGTGATCCCGCCGCTCGCCCTCCCAAGCTGCTGGGATT	1856
	Db	62	GlyLeuLysLeuLeuAlaSerAspLeuProAlaSerAlaSerGlnSerAlaGlylle	81
	Qy	1855	ACAGGCGTGAGCCACCACGCCGCGCCCTCTCTCTCTTTTTTTTTTTTTT	1808
	Db	82	ThrGlyMetSerHisHisSerTrpProGluArgthrSerPhellePhe	97

RESULT 10  
US-09-227-357-193  
Sequence 193, Application US/09227357  
Patent No. 6342581  
GENERAL INFORMATION:  
APPLICANT: Fischer et al.  
TITLE OF INVENTION: 123 Human Secreted Proteins  
FILE REFERENCE: P201091  
CURRENT APPLICATION NUMBER: US/09/227,357  
CURRENT FILING DATE: 1999-01-08  
EARLIER APPLICATION NUMBER: PCT/US98/13684  
EARLIER FILING DATE: 1998-07-07  
EARLIER APPLICATION NUMBER: 60/051,926  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,793  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,925  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,929  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,803  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,732  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,931  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,932  
EARLIER FILING DATE: 1997-07-08



EARLIER APPLICATION NUMBER: 60/051,916  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,930  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,918  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,920  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,733  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,795  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,919  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,928  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/055,722  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,723  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,948  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,949  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,953  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,950  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,947  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,964  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/056,360  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,684  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,984  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,954  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/058,785  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,664  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,660  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,661  
EARLIER FILING DATE: 1997-09-12  
NUMBER OF SEQ ID NOS: 672  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 193  
LENGTH: 110  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (110)  
OTHER INFORMATION: xaa equals stop translation  
US-09-227-357-193

Alignment Scores:  
Pred. No.: 3.15e-22 Length: 110  
Score: 278.00 Matches: 61  
Percent Similarity: 76.04% Conservative: 12  
Best Local Similarity: 63.54% Mismatches: 16  
Query Match: 5.63% Indels: 7  
DB: 4 Gaps: 1

US-09-966-880A-7 (1-2818) x US-09-227-357-193 (1-110)

QY 2126 TTTTCTTGTGAGACAGACTGTGCTGTGCGCCAGGCTGAGTGCACCGCAC 2067  
||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||  
Db 21 PheValPheLeuArgLeuSerLeuLeu-ProLysLeuGluCysAsn----- 38

QY 2066 GATCTCAGCTTACTGCAACCTCCACCTCTGGTTCAGCGATTCTCTGCTCAGGCTC 2007  
::: ||||||||| ||||||||| ||||||||| ||||||||| |||||||||  
Db 39 -LeuGlySer--LeuGlnProProProArgPheGlnArgPheSerCysLeuSerLeu 57  
QY 2006 CCAAGTAGCTGGGATTACAGGTGCCTGCTACACGCTGGCTAATTTTGCATTTTGA 1947  
::: ||||||||| ||| ||||||||| ||||||||| ||||||||| |||||||||  
Db 58 LeuAsnSerTrpAspTrpArgProProHisLeuAlaAsnPheCysValValSer 77  
QY 1946 ACAGAGGGGTTTGGCCATGTTGGCCAGGCTGCTCAAACTCCTGACCACAGTGATCC 1887  
||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||  
Db 78 ArgGlyGlyValSerSerCysTrpProGlyTrpSerArgThrProAsp--LeuMetIle 97  
QY 1886 GCCCGGCTCGGCTCCCAAGTGTGGGATTACAGCG 1849  
|| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||  
Db 97 rgLeuProArgProProArgValLeuGlyLeuGlnAla 109

RESULT 11

US-08-687-895-1  
; Sequence 1, Application US/08687895  
; Patent No. 5747319  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Hawkins, Phillip R.  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; CURRENT APPLICATION DATA: US/08/687,895  
; APPLICATION NUMBER: US/08/687,895  
; FILING DATE: Filed Herewith  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0109 US  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 222 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY: MUSCNOT1  
; CLONE: 57953  
US-08-687-895-1

Alignment Scores:  
Pred. No.: 2.81e-21 Length: 222  
Score: 271.00 Matches: 63  
Percent Similarity: 50.00% Conservative: 25  
Best Local Similarity: 35.80% Mismatches: 76  
Query Match: 5.48% Indels: 12  
DB: 1 Gaps: 5

US-09-966-880A-7 (1-2818) x US-08-687-895-1 (1-222)

QY 110 TTTCTTTACCAATTCAAAATGTCGCTGGCTAAGGTCGGCTGAGACTACCTGTC 169  
||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||  
||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |||||||||

Db 52 PheLysPheGlnPheArgAsnValGluTyrSerSerGlyArgAsnLysThrPheLeuCys 71  
Qy 170 TACGTAGTGAAGGCGGTGACAGTGTACATCTCTTTTCACTGGCATTGGTTACTTTCGC 229  
Db 72 TyrValValGlnGluGlyLysGlyGlnValGlnAlaSerArgGlyTyrLeuGlu 91  
Qy 230 AATAAGAAC--GGCTCCAGGTGAATGCTCTCTCCGCTACATCTCGGACTGGAC 286  
Db 92 AspGluHisAlaAlaAlaHisAlaGluAlaPheAsnThrIleLeuPro--Ala 110  
Qy 287 CTAGACCTGGCGCTCTACCGGTCACCTGGTTCACCTCTCGGACCCCTGCTACGAC 346  
Db 111 PheAspProAlaLeuArgTyrAsnValThrTyrValSerSerProCysAlaAla 130  
Qy 347 TGTCCCGCAGATGTGCGCGACTTCTCGAGGGAACCCCACTCTGAGGATCTTC 406  
Db 131 CysAlaAspArgIle\*\*LysThrLeuSerLysThrLysAsnLeuArgLeuLeuLeu 150  
Qy 407 ACCGCGGCTCTACTTCTGTGAGGACCGCAAGCTGAGCCCGAG-----GGGCTG 457  
Db 151 ValGlyArgLeuPheMetTrpGlu-----GluProGluIleGlnAlaAlaLeu 166  
Qy 458 CGGCGGCTGCACCGCGCGGTGCAATAGCCATCATGACCTTCAAGATTATTTTAC 517  
Db 167 LysLysLeuLysGluAlaGlyCysLysLeuArgIleMetLysProGlnAspPheGluTyr 186  
Qy 518 TGTGGTAATCTTTGTAGAAACCATGAA-----AGAACTTTCAAGCCTGGGAA 568  
Db 187 ValTrpGlnAsnPheValGlnGluGluGluGluGluGluGluGluGluGluGlu 206  
Qy 569 GGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGGCGATC 616  
Db 207 AspIleGlnGluAsnPheLeuTyrTyrGluGluLysLeuAlaAspIle 222

## RESULT 12

US-09-040-482-1  
; Sequence 1, Application US/09040482  
; Patent No. 5916556  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Hawkins, Phillip R.  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq Version 1.5  
; APPLICATION DATA:  
; APPLICATION NUMBER: US/09/040,482  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/687,895  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0109 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 222 amino acids  
; TYPE: amino acid

; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY: MUSCNOT1  
; CLONE: 57953  
; US-09-040-482-1

Alignment Scores:  
Pred. No.: 2,81e-21 Length: 222  
Score: 271.00 Matches: 63  
Percent Similarity: 50.00% Conservative: 25  
Best Local Similarity: 35.80% Mismatches: 76  
Query Match: 5.48% Indels: 12  
DB: 2 Gaps: 5

US-09-966-880A-7 (1-2818) x US-09-040-482-1 (1-222)

Qy 110 TTTCTTTTACCAATTCAAAATGTCCGCTGGGCTAAGGTCGGCTGAGACCTACCTGTGC 169  
Db 52 PheLysPheGlnPheArgAsnValGluTyrSerSerGlyArgAsnLysThrPheLeuCys 71  
Qy 170 TACGTAGTGAAGGCGGTGACAGTGTACATCTCTTTTCACTGGACTTTGGTTACTTTCGC 229  
Db 72 TyrValValGlnGluGlyLysGlyGlnValGlnAlaSerArgGlyTyrLeuGlu 91  
Qy 230 AATAAGAAC--GGCTCCACACGTGGAATTCCTCTCTCCGCTACATCTCGGACTGGAC 286  
Db 92 AspGluHisAlaAlaAlaHisAlaGluAlaPheAsnThrIleLeuPro--Ala 110  
Qy 287 CTAGACCTGGCGCTCTACCGGTCACCTGGTTCACCTCTCGGACCCCTGCTACGAC 346  
Db 111 PheAspProAlaLeuArgTyrAsnValThrTyrValSerSerProCysAlaAla 130  
Qy 347 TGTCCCGCAGATGTGCGCGACTTCTCGAGGGAACCCCACTCTGAGGATCTTC 406  
Db 131 CysAlaAspArgIle\*\*LysThrLeuSerLysThrLysAsnLeuArgLeuLeuLeu 150  
Qy 407 ACCGCGGCTCTACTTCTGTGAGGACCGCAAGCTGAGCCCGAG-----GGGCTG 457  
Db 151 ValGlyArgLeuPheMetTrpGlu-----GluProGluIleGlnAlaAlaLeu 166  
Qy 458 CGGCGGCTGCACCGCGCGGTGCAATAGCCATCATGACCTTCAAGATTATTTTAC 517  
Db 167 LysLysLeuLysGluAlaGlyCysLysLeuArgIleMetLysProGlnAspPheGluTyr 186  
Qy 518 TGTGGTAATCTTTGTAGAAACCATGAA-----AGAACTTTCAAGCCTGGGAA 568  
Db 187 ValTrpGlnAsnPheValGlnGluGluGluGluGluGluGluGluGluGluGlu 206  
Qy 569 GGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGGCGATC 616  
Db 207 AspIleGlnGluAsnPheLeuTyrTyrGluGluLysLeuAlaAspIle 222

## RESULT 13

US-09-227-357-285  
; Sequence 285, Application US/09227357  
; Patent No. 6342581  
; GENERAL INFORMATION:  
; APPLICANT: Fischer et al.  
; TITLE OF INVENTION: 123 Human Secreted Proteins  
; FILE REFERENCE: P2010P1  
; CURRENT APPLICATION NUMBER: US/09/227,357  
; CURRENT FILING DATE: 1999-01-08  
; EARLIER APPLICATION NUMBER: PCT/US98/13684  
; EARLIER FILING DATE: 1998-07-07  
; EARLIER APPLICATION NUMBER: 60/051,926  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/052,793  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,925  
; EARLIER FILING DATE: 1997-07-08  
; EARLIER APPLICATION NUMBER: 60/051,929

EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,803  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,732  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,931  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,932  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,916  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,930  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,918  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,920  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,733  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/052,795  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,919  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/051,928  
EARLIER FILING DATE: 1997-07-08  
EARLIER APPLICATION NUMBER: 60/055,722  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,723  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,948  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,949  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,953  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,950  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,947  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,964  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/056,360  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,684  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,984  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/055,954  
EARLIER FILING DATE: 1997-08-18  
EARLIER APPLICATION NUMBER: 60/058,785  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,664  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,660  
EARLIER FILING DATE: 1997-09-12  
EARLIER APPLICATION NUMBER: 60/058,661  
EARLIER FILING DATE: 1997-09-12  
NUMBER OF SEQ ID NOS: 672  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 285  
LENGTH: 122  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-227-357-285

Alignment Scores:  
Pred. No.: 7,19e-19 Length: 122  
Score: 248.00 Matches: 56  
Percent Similarity: 61.86% Conservative: 4  
Best Local Similarity: 57.73% Mismatches: 36  
Query Match: 5.02% Indels: 1  
DB: 4 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-227-357-285 (1-122)  
QY 2054 CTGCAACCTCCACCTCCTGGTTCAAGCGATTCTCTGCTCAGCTCCAGCTCCAGTAGCTGG 1995  
Db : : : : :  
1 MetGlnAlaLeuProGlyPheLysGlnPheSerCysLeuSerLeuProSerArgTTP 20  
QY 1994 GATTACAGGTGCTGTACACGCTGGCTAATTTTGGATTTTGGATTTGAGTACAGACGGGGTT 1935  
Db : : : : :  
21 AspTyrGlyCysAlaThrGlnHisProAlaAsnPheCysIlePheArgArgAspArgVal 40  
QY 1934 TTGCCATGTTGCCAGGCTGCTCAAACTCCTGACACAGGTGATCCGCCGCTCGGC 1875  
Db : : : : :  
41 -SerHisValGlyGlnAlaGlyLeuLysLeuLeuThrSerValAspProProAlaTrpAl 60  
QY 1874 CTCCCAAAGTGTGGATTACAGCGTGAGCCACACCGCCGCTCTCTCTTTCTTTT 1815  
Db : : : : :  
60 aSerGlnSerAlaGlyIleThrGlyLysSerHisCysAlaGlnLeuHisCysCysPh 80  
QY 1814 TTTTFTTTTGTGAGACAGGATCTTGTATGTGTCGCCAGGCTG 1766  
Db : : : : :  
80 eLeuLeuValLysArgAspGlnProLeuGluLysCysLeuArgLeu 96

RESULT 14

US-08-687-895-3  
Sequence 3, Application US/08687895  
Patent No. 5747319  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Hawkins, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,895  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0109 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 116 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 436941  
US-08-687-895-3

Alignment Scores:  
Pred. No.: 1,51e-18 Length: 116  
Score: 245.00 Matches: 51  
Percent Similarity: 61.82% Conservative: 17  
Best Local Similarity: 46.36% Mismatches: 32  
Query Match: 4.95% Indels: 10  
DB: 1 Gaps: 3

US-09-966-880A-7 (1-2818) x US-08-687-895-3 (1-116)  
QY 305 TACCGCGTCCACCTGGTTCACCTCTCGAGCCCTGCTACGAC-----TGTGCCCGACAT 358  
Db 7 TyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGlyGlu 26  
QY 359 GTGGCCGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCCCTC 418  
Db 27 ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgIle 46  
QY 419 TACTTCTGTGAGACCGCAAGGCTGAGCCC-----GAGGGGCTCGCGCGCTGCAC 469  
Db 47 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 61  
QY 470 CGCGCGGGGTGCAATAGCCATGACCTTCAAGATTTATTTTACTGCTGGAATACT 529  
Db 62 AspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheGluTyrCysTrpAspThr 81  
QY 530 TTTGTAGAAAACCATGAAAGAACTTTCAAAGCTGGGAAGGCTGCATGAAAATTCAGTT 589  
Db 82 PheValTyrArgGlnGlyCysProPheGlnProTrpAspGlyLeuGluGluHisSerGln 101  
QY 590 CGTCTCTCCAGACAGCTTCGGCGCATCCTT 619  
Db 102 AlaLeuSerGlyArgLeuArgAlaIleLeu 111

RESULT 15  
US-08-816-241-3  
; Sequence 3, Application US/08816241  
; Patent No. 5804185  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surva K.  
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/816,241  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0239 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 116 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 436941  
; US-08-816-241-3

Alignment Scores:  
Pred. No.: 1-51e-18 Length: 116  
Score: 245.00 Matches: 51  
Percent Similarity: 61.82% Conservative: 17  
Best Local Similarity: 46.36% Mismatches: 32  
Query Match: 4.95% Indels: 10  
DB: 1 Gaps: 3  
US-09-966-880A-7 (1-2818) x US-08-816-241-3 (1-116)  
QY 305 TACCGCGTCCACCTGGTTCACCTCTCGAGCCCTGCTACGAC-----TGTGCCCGACAT 358  
Db 7 TyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGlyGlu 26  
QY 359 GTGGCCGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCCCTC 418  
Db 27 ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgIle 46  
QY 419 TACTTCTGTGAGACCGCAAGGCTGAGCCC-----GAGGGGCTCGCGCGCTGCAC 469  
Db 47 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 61  
QY 470 CGCGCGGGGTGCAATAGCCATGACCTTCAAGATTTATTTTACTGCTGGAATACT 529  
Db 62 AspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheGluTyrCysTrpAspThr 81  
QY 530 TTTGTAGAAAACCATGAAAGAACTTTCAAAGCTGGGAAGGCTGCATGAAAATTCAGTT 589  
Db 82 PheValTyrArgGlnGlyCysProPheGlnProTrpAspGlyLeuGluGluHisSerGln 101  
QY 590 CGTCTCTCCAGACAGCTTCGGCGCATCCTT 619  
Db 102 AlaLeuSerGlyArgLeuArgAlaIleLeu 111

Search completed: June 14, 2003, 18:59:19  
Job time : 37.1417 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 14, 2003, 18:26:00 ; Search time 90.7352 Seconds

(without alignments)  
6641.044 Million cell updates/sec

Title: US-09-966-880A-7

Perfect score: 4948

Sequence: 1 agagaccatcatattga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delopt 6.0, Delext 7.0

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 817286

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp

-Q/cgn2\_1/USPTO.spool/US09966880/runat\_14062003\_175526\_10413/app.query.fasta\_1.9493

-DB=PublishedApplications\_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1

-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62

-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100

-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0

-MAXLEN=2000000000 -USER=US09966880 @CGN\_1\_1.105.@runat\_14062003\_175526\_10413

-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100

-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5

-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications\_AA:

1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pap.\*

2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pap.\*

3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pap.\*

4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pap.\*

5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pap.\*

6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pap.\*

7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pap.\*

8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pap.\*

9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pap.\*

10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pap.\*

11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pap.\*

12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pap.\*

13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pap.\*

14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	21.9	198	9	US-09-966-880A-8
2	1008	20.4	198	9	US-09-966-880A-2
3	390	7.9	384	10	US-09-729-674-174
4	381.5	7.7	239	10	US-09-800-729-193

381.5	7.7	310	10	US-09-800-729-190	Sequence 190, App
369.5	7.5	222	10	US-09-925-300-1639	Sequence 1639, Ap
365.5	7.4	109	9	US-09-989-919-105	Sequence 105, App
365	7.4	94	9	US-10-001-835-135	Sequence 135, App
361	7.3	107	9	US-09-989-920-233	Sequence 233, App
359	7.3	144	9	US-10-011-585A-145	Sequence 145, App
358	7.3	100	9	US-10-016-157A-171	Sequence 171, App
351.5	7.1	126	9	US-10-082-830-207	Sequence 207, App
351	7.1	172	10	US-09-764-864-1058	Sequence 1058, Ap
349.5	7.1	121	9	US-09-989-920-198	Sequence 198, App
343	6.9	198	9	US-09-986-480-433	Sequence 433, App
338.5	6.9	123	9	US-09-764-868-888	Sequence 888, App
338.5	6.9	170	9	US-09-986-480-282	Sequence 282, App
332	6.7	115	10	US-09-995-494-79	Sequence 79, Appl
332	6.7	132	9	US-10-012-896-573	Sequence 573, App
332	6.7	132	9	US-09-895-793-573	Sequence 573, App
332	6.7	132	9	US-09-895-814-573	Sequence 573, App
332	6.7	132	10	US-09-759-143-573	Sequence 573, App
332	6.7	132	10	US-09-780-669-573	Sequence 573, App
332	6.7	132	10	US-09-822-827-573	Sequence 573, App
332	6.7	135	9	US-10-012-896-884	Sequence 884, App
332	6.7	135	9	US-09-895-793-884	Sequence 884, App
332	6.7	135	9	US-09-895-814-884	Sequence 884, App
332	6.7	135	10	US-09-759-143-884	Sequence 884, App
332	6.7	135	10	US-09-780-669-884	Sequence 884, App
332	6.7	135	10	US-09-822-827-884	Sequence 884, App
328.5	6.7	122	9	US-10-074-475-190	Sequence 190, App
320.5	6.5	118	10	US-09-764-898-157	Sequence 157, App
319.5	6.5	117	9	US-10-001-835-203	Sequence 203, App
315.5	6.4	121	10	US-09-764-877-1641	Sequence 1641, Ap
312.5	6.3	109	9	US-10-001-835-154	Sequence 154, App
311	6.3	90	9	US-10-001-873-34	Sequence 34, Appl
310	6.3	93	9	US-09-764-891-5337	Sequence 5337, Ap
309	6.3	86	9	US-10-011-585A-154	Sequence 154, App
309	6.3	361	10	US-09-995-494-107	Sequence 107, App
308	6.2	157	9	US-10-001-835-197	Sequence 197, App
304.5	6.2	100	9	US-10-016-157A-187	Sequence 187, App
303	6.1	375	9	US-10-146-130-2	Sequence 2, Appli
303	6.1	375	9	US-09-964-667-2	Sequence 2, Appli
303	6.1	375	9	US-09-872-968-2	Sequence 2, Appli
303	6.1	375	9	US-10-092-934-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1  
US-09-966-880A-8  
; Sequence 8, Application US/09966880A  
; Patent No. US2002016473A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966,880A  
; CURRENT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 198  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-966-880A-8

Alignment Scores:

Pred. No.: 1.43e-102 Length: 198  
 Score: 1086.00 Matches: 198  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 21.95% Indels: 0  
 DB: 9 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-966-880A-8 (1-198)

QY 80 ATGACAGCCCTCTGTAGTGAACCGAGGAGTCTTTTACCAATTCAAAATGTCGCTGG 139  
 Db 1 MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
 QY 140 GCTAAGGTCGCGTGGAGCTTACCTGTCTAGTGAAGAGGCGTGCACGTGCTACA 199  
 Db 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40  
 QY 200 TCCTTTTCACTGGACTTTCGCTATCTCCCAATAGAAGCGGTGCACGTGGAATGCTC 259  
 Db 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60  
 QY 260 TTCCTCCGCTACATTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGTGCACCTGG 319  
 Db 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
 QY 320 TTCACCTCTCGAGCCCTCTGCTACGACTGTGCGCCGCTCTACTTCTGTGAGGACCGCAAG 379  
 Db 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
 QY 380 ACCCCCACTCAGTCTGAGGATCTTACCGCGGCTCTACTTCTGTGAGGACCGCAAG 439  
 Db 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
 QY 440 GCTGAGCCGAGGGCTCGCGGCTGACGACTGTGCGCGGCTCTACTTCTGTGAGGACCGCAAG 499  
 Db 121 AlaGluProGluGlyLeuArgLysHisArgAlaGlyValGlnIleAlaIleMetThr 140  
 QY 500 TTCAAAGATTATTTTACTGCTGAATACTTTTGTAGAAAACCATGAAGAACTTTCAAA 559  
 Db 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
 QY 560 GCCTGGGAGGCTGCATGAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCCCT 619  
 Db 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
 QY 620 TTGCCCTGTATGAGTTGATGACTTACGAGAGCGCATTTCTGCTACTTTGGGACTT 673  
 Db 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198

RESULT 2

US-09-966-880A-2  
 ; Sequence 2, Application US/09966880A  
 ; Patent No. US20020164743A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Honjo, Tasuku  
 ; APPLICANT: Muramatsu, Masamichi  
 ; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
 ; FILE REFERENCE: 06501-088001  
 ; CURRENT APPLICATION NUMBER: US/09/966,880A  
 ; CURRENT FILING DATE: 2001-09-28  
 ; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
 ; PRIOR FILING DATE: 2000-03-28  
 ; PRIOR APPLICATION NUMBER: JP 11-371382  
 ; PRIOR FILING DATE: 1999-12-27  
 ; PRIOR APPLICATION NUMBER: JP 11-178999  
 ; PRIOR FILING DATE: 1999-06-24  
 ; PRIOR APPLICATION NUMBER: JP 11-87192  
 ; PRIOR FILING DATE: 1999-03-29  
 ; NUMBER OF SEQ ID NOS: 36  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 198  
 ; TYPE: PRT

ORGANISM: Mus musculus

US-09-966-880A-2

Alignment Scores:

Pred. No.: 1.51e-94 Length: 198  
 Score: 1008.00 Matches: 183  
 Percent Similarity: 95.94% Conservative: 6  
 Best Local Similarity: 92.89% Mismatches: 8  
 Query Match: 20.37% Indels: 0  
 DB: 9 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-966-880A-2 (1-198)

QY 80 ATGACAGCCCTCTGTAGTGAACCGAGGAGTCTTTTACCAATTCAAAATGTCGCTGG 139  
 Db 1 MetAspSerLeuLeuMetLysGlnLysLysPheLeuTyrHisPheLysAsnValArgTrp 20  
 QY 140 GCTAAGGTCGCGTGGAGCTTACCTGTCTAGTGAAGAGGCGTGCACGTGCTACA 199  
 Db 21 AlaLysGlyArgHisGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
 QY 200 TCCTTTTCACTGGACTTTCGCTATCTCCCAATAGAAGCGGTGCACGTGGAATGCTC 259  
 Db 41 SerCysSerLeuAspPheGlyHisLeuArgAsnLysSerGlyCysHisValGluLeu 60  
 QY 260 TTCCTCCGCTACATTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGTGCACCTGG 319  
 Db 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
 QY 320 TTCACCTCTCGAGCCCTCTGCTACGACTGTGCGCCGCTCTACTTCTGTGAGGACCG 379  
 Db 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaGluPheLeuArgTrp 100  
 QY 380 ACCCCCACTCAGTCTGAGGATCTTACCGCGGCTCTACTTCTGTGAGGACCGCAAG 439  
 Db 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
 QY 440 GCTGAGCCGAGGGCTCGCGGCTGACGACTGTGCGCGGCTCTACTTCTGTGAGGACCGCAAG 499  
 Db 121 AlaGluProGluGlyLeuArgLysHisArgAlaGlyValGlnIleGlyIleMetThr 140  
 QY 500 TTCAAAGATTATTTTACTGCTGAATACTTTTGTAGAAAACCATGAAGAACTTTCAAA 559  
 Db 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnArgGluArgThrPheLys 160  
 QY 560 GCCTGGGAGGCTGCATGAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCCCT 619  
 Db 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuThrArgGlnLeuArgArgIleLeu 180  
 QY 620 TTGCCCTGTATGAGTTGATGACTTACGAGAGCGCATTTCTGCTACTTTGGGCA 670  
 Db 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgMetLeuGly 197

RESULT 3

US-09-729-674-174  
 ; Sequence 174, Application US/09729674  
 ; Patent No. US20010039335A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Jacobs, Kenneth  
 ; APPLICANT: McCoy, John M.  
 ; APPLICANT: Lavallie, Edward R.  
 ; APPLICANT: Collins-Racie, Lisa A.  
 ; APPLICANT: Evans, Cheryl  
 ; APPLICANT: Merberg, David  
 ; APPLICANT: Treacy, Maurice  
 ; APPLICANT: Agostino, Michael J.  
 ; APPLICANT: Steininger II, Robert J.  
 ; APPLICANT: Spaulding, Vikki  
 ; APPLICANT: Wong, Gordon G.  
 ; APPLICANT: Clark, Hilary  
 ; APPLICANT: Fechtel, Kim  
 ; APPLICANT: Genetics Institute, Inc.  
 ; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM



QY	170	TACGTAAGTGAAG-----AGCGCTGCACAGTGCTATACCTCTTTTCACGTGGACTTTGGTTAT	223
Db	69	PheThrValcIuglylellysArgSerValValserTrpLysThr-----GlyVal	86
QY	224	CUTTGGCAAT-----AAGACGGCTGCCACCTGGAAATTGCTCTTTCCCTCCGCTAC	271
Db	87	PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp	106
QY	272	ATCTCGGACTGGACCTAGACCCTGGCGCTGTACCGCTCACCTGGTTACCTCCTCGG	331
Db	107	PheCysAspAspIleLeuSerProAsnThrLysTyrgInValTrpTrpTrpSerTrp	126
QY	332	AGCCCTGCTACGACTGTGGCCGACATGTGGCGACTTTCTCGAGGGAACCCCNAACCTC	391
Db	127	SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal	146
QY	392	AGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGCGTAGCCCCGAG	451
Db	147	AsnLeuThrIlePheThrAlaArgLeuTyrrPheGln---TyrProCysTyrgInGlu	165
QY	452	GGCTGCGCGCGCTGCACGCGCGCGGGTGCAAATAGCCATCATGACCTTCAAAGATTAT	511
Db	166	GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrgLuAspPhe	185
QY	512	TTTTACTGCTGGAACTACTTTGTAGAAAACCATGAAGAATCTTTCAAGCTCGGAAGG	571
Db	186	LysTyrcysTrpGluAsnPheValtyrAsnAspAsnGluProPheLysProTrpLysGly	205
QY	572	CTGCATGAAAAATTACGTTCTCTCCACAGACGCTTCGGCGCATCCTTT	619
Db	206	LeuLysThrAsnPheArgLeuLeuLysArgLeuArgGluSerLeu	22

```

RESULT 7
US-09-989-919-105
; Sequence 105, Application US/09989919
; Patent No. US20020164344A1
; GENERAL INFORMATION:
; APPLICANT: Macina, Roberto
; APPLICANT: Recipon, Hervé
; APPLICANT: Pluta, Jason
; APPLICANT: Ghosh, Malavika
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Colon Specific Genes and
; FILE REFERENCE: DEX-0289
; CURRENT APPLICATION NUMBER: US/09/989,919
; CURRENT FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: 60/252,505
; PRIOR FILING DATE: 2000-11-22
; NUMBER OF SEQ ID NOS: 124
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 105
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-989-919-105

Alignment Scores:
Pred. No.:
Score: 1.29e-28
Percent Similarity: 365.50
Best Local Similarity: 80.56%
Query Match: 74.07%
DB: 19
      2
      Gaps: 1

US-09-966-880A-7 (1-2818) x US-09-989-919-105 (1-109)
QY      2126 TTTTTHTTTTTTTGAGACAGAGTCTTGCTCTTGCGCCAGGCTGGAGTCGAACGGCAC 2067
Db      1 PhePhePheSerLeuArgGlnSer-LeuLeuLeuLeuProArgLeuGluPheAsnGlyTh 20
QY      2066 GATCTCAGCTTACTGCAACCTCCACCTCCTGGTTCACAGCGATTCTCTGGCTCAGCGCTC 2007

```



Db 20 rtleLeuAlaTyHisAsnLeuCysLeuLeuGlySerSerAsnSerProAlaSerGlyse 40  
QY 2006 CCAAGTAGCTGGATTACAGGTGCCTGCTACACCGCTGGCTAAATTTTTCATTTTGACT 1947  
Db 40 rGlnValAlaGlyIleThrGlyMetCysHisThrArgLeuIlePheValPheLeuVa 60  
QY 1946 ACAGACGGGGTTTTCATGTTGGCCAGGCTGTCTCAAACTCCTGACACAGGTGATCC 1887  
Db 60 lGluThrGlyTyLeuHisValGlyGlnAlaGlyLeuGluLeuLeuThrSerGlyAspPr 80  
QY 1886 GCCCGCTGGCTCCAAAGTCTGGGATTACAGCGGTGAGCCACCGCCGGGCC-- 1829  
Db 80 oProThrSerAlaSerGlnSerAlaGlyIleThrGlyValSerArgHisAlaTrpProSe 100  
QY 1828 -TCTCTCTTCTTTTTTTTTTT 1808  
Db 100 rSerAlaPheIleHisIlePhe 107

RESULT 8

US-10-001-835-135  
; Sequence 135, Application US/10001835  
; Patent No. US20020160387A1

GENERAL INFORMATION:

; APPLICANT: Salceda, Susana  
; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Herve  
; APPLICANT: Cafferey, Robert  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua

; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and Pro

; FILE REFERENCE: DEX-0277

; CURRENT APPLICATION NUMBER: US/10/001,835

; PRIOR FILING DATE: 2001-11-20

; PRIOR APPLICATION NUMBER: 60/249,997

; NUMBER OF SEQ ID NOS: 228

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 135

; LENGTH: 94

; TYPE: PRT

; ORGANISM: Homo sapien

US-10-001-835-135

Alignment Scores:  
Pred. No.: 1,36e-28 Length: 94  
Score: 365.00 Matches: 71  
Percent Similarity: 84.78% Conservative: 7  
Best Local Similarity: 77.17% Mismatches: 14  
Query Match: 7.39% Indels: 0  
DB: 9 Gaps: 0

US-09-966-880A-7 (1-2818) x US-10-001-835-135 (1-94)

QY 2101 TTGCTCTTTCGCCAGGTGGAGTGCACCGCACGATCTCAGCTTACTGCAACCTCCAC 2042  
Db 3 LeuThrLeuSerProArgLeuGluCysSerGlyThrValSerAlaHisCysAsnLeuHis 22  
QY 2041 CTCCTGGGTCAAGCATTTCTCTGCTCAGCTCCCAAGTAGCTGGGATTACAGTGCC 1982  
Db 23 LeuLeuGlySerSerAspSerProAlaSerValSerAlaValAlaGlyThrThrGlyVal 42  
QY 1981 TGCTACACGCTGGCTAATTTTTCATTTTTCAGTACACAGCGGGTTTTCATTTTTCG 1922  
Db 43 ArgHisHisAlaTrpLeuIlePheIlePheLeuValGluThrValPheCysHisValGly 62  
QY 1921 CAGGTGGTCTCAAACTCCTGACACAGGTGATCCCGCGGTGCTGGCCTCCCAAAAGTCT 1862  
Db 63 GlnAlaGlyLeuLysLeuLeuThrSerGlyAspProProThrSerAlaSerAlaSerThr 82  
QY 1861 GGGATTACAGCGGTGAGCCACACCGCCCGCCCTCT 1826  
Db 83 GlyIleThrGlyMetSerHisCysAlaTrpProSer 94

RESULT 9

US-09-989-920-233  
; Sequence 233, Application US/0989920  
; Patent No. US20020172957A1

GENERAL INFORMATION:

; APPLICANT: Macina, Roberto  
; APPLICANT: Recipon, Herve  
; APPLICANT: Chen, Sei-Yu  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua

; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and P

; FILE REFERENCE: DEX-0291

; CURRENT APPLICATION NUMBER: US/09/989,920

; CURRENT FILING DATE: 2001-11-21

; PRIOR APPLICATION NUMBER: 60/252,500

; PRIOR FILING DATE: 2000-11-22

; NUMBER OF SEQ ID NOS: 284

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 233

; LENGTH: 107

; TYPE: PRT

; ORGANISM: Homo sapien

US-09-989-920-233

Alignment Scores:

Pred. No.: 3.7e-28 Length: 107  
Score: 361.00 Matches: 76  
Percent Similarity: 79.25% Conservative: 18  
Best Local Similarity: 71.70% Mismatches: 19  
Query Match: 7.31% Indels: 3  
DB: 9 Gaps: 1

US-09-966-880A-7 (1-2818) x US-09-989-920-233 (1-107)

QY 2129 TTTTTCCTTTTCTGAGA-----CAGAGTCCTTCTCTTCTGCTCCGCCAGGTGGAGTG 2076  
Db 1 PhePhePhePheLeuLeuLeuPheCysAspSerLeuAla-LeuSerProArgLeuGlnCy 20  
QY 2075 CAACGGCAGATCTCAGCTTACTGCAACCTCCACCTCCCTGGGTTCAGCGATTCTCTG 2016  
Db 20 sSerGlyThrIleSerAlaHisCysAsnLeuLeuGlySerSerAsnSerProva 40  
QY 2015 CTCAGCTCCCAAGTAGCTGAGTTCAGTGTCTACAGCTGCTACACGCTGGCTAATTTTTC 1956  
Db 40 lSerAlaSerTrpValAlaGlyThrThrGlyAlaCysHisHisAlaTrpLeuThrPheVa 60  
QY 1955 ATTTGAGTACAGACGGGGTTTTCCTATGTTGCCAGGTGGTCTCAAACTCCTGACCAC 1896  
Db 60 lPheLeuValGluThrGlyPheHisHisValGlyGlnAlaGlyLeuGluPheLeuThrSe 80  
QY 1895 AGGTGATCCCGCGGCTCGGCTCCCAAAAGTGTGGGATTACAGCGGTGAGCCACCGC 1836  
Db 80 rGlyAspProProAlaLeuAlaSerGlnSerAlaGluIleThrGlyValSerHisArgAl 100  
QY 1835 CCGGCCCTCTCTTT 1820  
Db 100 atrProValCysPhe 105

RESULT 10

US-10-011-585A-145  
; Sequence 145, Application US/10011585A  
; Publication No. US20030039986A1

GENERAL INFORMATION:

; APPLICANT: Sun, Yongming  
; APPLICANT: Recipon, Herve  
; APPLICANT: Chen, Sei-Yu  
; APPLICANT: Liu, Chenghua

; TITLE OF INVENTION: Compositions and Methods Relating to Prostate Specific

; FILE REFERENCE: DEX-0261

; CURRENT APPLICATION NUMBER: US/10/011,585A

; CURRENT FILING DATE: 2002-03-14



Db 29 LeuAlaMetSerGlnSer--IleMetAlaLeuSerProArgLeuGluCysAsnGlyAlaVal 48  
QY 2064 TCTCAGCTTACTGCAACCTCCACCTCTGGGTTCAAGGATCTCTCTGCTCAGCGCTCCC 2005  
:||||| :||||| ||| ||||||||||||||| ||||||| ||||||| |||||||  
Db 48 alserGlyHisCysAsnProCysLeuProGlySerSerAspSerProProSerAlaSerG 68  
QY 2004 AGTACTGGGATTACAGTGCTGCTACACGCCCTGGCTAAATTTTGGCATTTTGAGTAC 1945  
:||||| :||||| ||| :||||| ||||||| ||||||| ||| |||  
Db 68 InValAlaGlyIleThrGlyThrCysHisAlaArgLeuIlePheValPheLeuValG 88  
QY 1944 AGACGGGGTTTGGCCATGTGGCAGGCTGGTCTCAAACTCTGACCACAGGTGATCCGC 1885  
: : ||||| ||||||| ||||||| ||||||| :||||| :||||| :|||||  
Db 88 luMetGlyPheHisValGlyGlnAlaGlyLeuGluLeuThrSerGlyAspLeuP 108  
QY 1884 CCGGCTCGGCCCTCCCAAGTGTGGGATTACAGGCTGAGCCACACCGCCGCCGCC 1829  
|| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 108 roThrSerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArgPro 126

## RESULT 13

US-09-764-864-1058  
; Sequence 1058, Application US/09764864  
; Patent No. US20020132753A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT223  
; CURRENT APPLICATION NUMBER: US/09/764,864  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 1792  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1058  
; LENGTH: 172  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-764-864-1058

Alignment Scores:  
Pred. No.: 4,8e-27 Length: 172  
Score: 351.00 Matches: 69  
Percent Similarity: 80.00% Conservative: 7  
Best Local Similarity: 72.63% Mismatches: 19  
Query Match: 7.11% Indels: 1  
DB: 10 Gaps: 0

US-09-966-880A-7 (1-2818) x US-09-764-864-1058 (1-172)

QY 2110 GACAGAGTCTTGCTCTTGTCGCCAGGCTGGAGTGCACCGCAGCATCTCAGCTTACTGC 2051  
||| ||| ||||||||||||||| ||||||| :||||| :||||| :|||||  
Db 1 AspGluValSerLeuLeuSerProArgLeuValCysAsnGlyMetMetSerAlaHisCys 20  
QY 2050 AACCTCCACCTCTCTGGGTTCAAGCGATTCTCTGCTCAGCCTCCCAAGTAGCTGGGATT 1991  
||| ||| ||||||||||||||| ||||||| ||||||| ||||||| |||||||  
Db 21 LysLeuArgLeuProGlySerSerAspSerProThrSerAlaSerGlnValAlaGlyIle 40  
QY 1990 ACAGGTGCCTTACCACGCCCTGGCTAATTTTGGCATTTTGATACAGCGGGTTTTCG 1931  
||||| ||| :||||| ||||||| ||||||| ||| :||| :||| :|||  
Db 41 ThrGlyThrCysHisAlaTrpLeuIlePheValLeuLeuValGluMetGlyTyrAsp 60  
QY 1930 CATGTTGGCAGGCTGGTCTCAAACTCTGACACAGGTATCCGCCCGG-CTCGGCCCTC 1872  
:||||| :||||| ||||||| ||||||| :||||| :||||| :|||||  
Db 61 HisLeuGlyGlnAlaAlaLeuLysLeuLeuThrSerGlyAspProThrSerGlyLeu 80  
QY 1871 CCAAACTGCTGGGATTACAGCGGTGAGCCACCGCCCGCCCTC 1827  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 81 ProLysCysTrpAspTyrArgHisGluProProLeuProAlaLeu 95

## RESULT 14

US-09-989-920-198  
; Sequence 198, Application US/09989920  
; Patent No. US20020172957A1  
; GENERAL INFORMATION:  
; APPLICANT: Macina, Roberto

; APPLICANT: Recipon, Herve  
; APPLICANT: Chen, Sei-yu  
; APPLICANT: Sun, Yongming  
; APPLICANT: Liu, Chenghua  
; TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific Genes and  
; FILE REFERENCE: DEX-0291  
; CURRENT APPLICATION NUMBER: US/09/989,920  
; CURRENT FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: 60/252,500  
; PRIOR FILING DATE: 2000-11-22  
; NUMBER OF SEQ ID NOS: 284  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 198  
; LENGTH: 121  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-989-920-198

## Alignment Scores:

Pred. No.: 5,92e-27 Length: 121  
Score: 349.50 Matches: 78  
Percent Similarity: 68.46% Conservative: 11  
Best Local Similarity: 60.00% Mismatches: 32  
Query Match: 7.08% Indels: 9  
DB: 9 Gaps: 3

US-09-966-880A-7 (1-2818) x US-09-989-920-198 (1-121)

QY 2119 TTTTCTGACAGAGTCTTGCTTGTGCCCCAGGCTGGAGTGCAACGGCAGCATCTCA 2060  
||||| ||| ||| ||||||| ||||||| :||||| :||||| :|||||  
Db 1 PhePheAlaAspGluValSerArgLeuSerProGlyLeuGluCysSerGlyValIleSer 20  
QY 2059 GCTTACTGCAACCTCCACCTCTGGGTTCAAGCGATTCTCTGCTCAGCCTCCCAAGTA 2000  
||||| :||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 21 AlaHisCysAsnPheHisLeuLeuGlySerSerSerProAlaSerAlaSerGlnVal 40  
QY 1999 GCTGGATTACAGGTGCTGTACACAGGCTGGCTAATTTTGGCATTTTGGATACAGAG 1940  
||| ||||| ||||| ||||| :||||| ||||| ||||| ||||| ||||| |||||  
Db 41 AlaGluIleThrGlyAlaCysHisProThrTrpLeuIlePheValIleLeuValGluThr 60  
QY 1939 GGGTTTGGCAGTGTGGCCAGGCTGTCTCAAACTCTGACACAGGTATCCGCCCGG 1880  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 61 GlyPheHisValGlyGlnAlaAsp---AlaLeuLeuThrSerGlyAspProProPhe 79  
QY 1879 TCGGCTCCCAAAGTCTGGGATTACAGCGGTACAGCGCCACCGCCCGCTCTCTCTT 1820  
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
Db 80 SerAlaSerGlnSerAlaGlyIleThrGlyValSerHisArgAlaArgProAla----- 97  
QY 1819 CTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 1760  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 98 -----AsnThrPheAlaLeuThrLeuGlyLeuLeuTyrlLysIleValMet 113  
QY 1759 AACACCTGTGTCAAGCCATCTCTCTCT 1730  
:||||| |||||  
Db 114 -----IleAlaMetGluValLeuProPro 121

## RESULT 15

US-09-986-480-433  
; Sequence 433, Application US/09986480  
; Publication No. US20030027999A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 143 Human Secreted Proteins  
; FILE REFERENCE: PS500P1  
; CURRENT APPLICATION NUMBER: US/09/986,480  
; CURRENT FILING DATE: 2001-11-08  
; PRIOR APPLICATION NUMBER: PCT/US00/12788  
; PRIOR FILING DATE: 2000-05-11  
; PRIOR APPLICATION NUMBER: US 60/134,068  
; PRIOR FILING DATE: 1999-05-13  
; NUMBER OF SEQ ID NOS: 456  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 433



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model  
Run on: June 18, 2003, 23:33:05 ; Search time 7195.7 Seconds  
(without alignments)  
11397.322 Million cell updates/sec  
Title: US-09-966-880A-7  
Perfect score: 2818  
Sequence: 1 adagaaccatcataattga.....aaaaaaaaaaaaaaaaaaaaa 2818

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0  
Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vl.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vl.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sv.\*
- 39: em\_higo\_hum.\*
- 40: em\_higo\_mus.\*
- 41: em\_higo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2791	99.0	2791	9	AB040431	AB040431 Homo sapi
2	2174.6	77.2	11204	9	AB040430	AB040430 Homo sapi
3	2174.6	77.2	71132	9	AC092184	AC092184 Homo sapi
4	1820	64.6	1837	9	BC006296	BC006296 Homo sapi
5	603.4	21.4	2440	10	AF132979	AF132979 Mus muscu
6	595.4	21.1	597	12	AF529828	AF529828 Mus muscu
7	594.4	21.1	596	9	AF529815	AF529815 Homo sapi
8	594.4	21.1	596	9	AF529816	AF529816 Homo sapi
9	594.4	21.1	596	9	AF529819	AF529819 Homo sapi
10	594.4	21.1	596	9	AF529820	AF529820 Homo sapi
11	594.4	21.1	596	9	AF529821	AF529821 Homo sapi
12	594.4	21.1	596	9	AF529822	AF529822 Homo sapi
13	594.4	21.1	596	9	AF529823	AF529823 Homo sapi
14	594.4	21.1	596	9	AF529824	AF529824 Homo sapi
15	594.4	21.1	596	9	AF529825	AF529825 Homo sapi
16	594.4	21.1	596	9	AF529826	AF529826 Homo sapi
17	594.4	21.1	596	9	AF529827	AF529827 Homo sapi
18	594.4	21.1	596	12	AF529829	AF529829 Mus muscu
19	594.4	21.1	596	12	AF529830	AF529830 Mus muscu
20	594.4	21.1	596	12	AF529831	AF529831 Mus muscu
21	594.4	21.1	596	12	AF529833	AF529833 Mus muscu
22	594.4	21.1	596	12	AF529835	AF529835 Mus muscu
23	594.4	21.1	596	12	AF529837	AF529837 Mus muscu
24	594.4	21.1	596	12	AF529839	AF529839 Mus muscu
25	593.4	21.1	595	12	AF529834	AF529834 Mus muscu
26	592.8	21.0	596	9	AF529817	AF529817 Homo sapi
27	592.8	21.0	596	9	AF529818	AF529818 Homo sapi
28	592.8	21.0	596	12	AF529832	AF529832 Mus muscu
29	592.8	21.0	596	12	AF529836	AF529836 Mus muscu
30	592.8	21.0	596	12	AF529838	AF529838 Mus muscu
31	592.2	21.0	597	12	AF529841	AF529841 Crictetu
32	592.2	21.0	597	12	AF529842	AF529842 Crictetu
33	592.2	21.0	597	12	AF529843	AF529843 Crictetu
34	592.2	21.0	597	12	AF529844	AF529844 Crictetu
35	592.2	21.0	597	12	AF529845	AF529845 Crictetu
36	592.2	21.0	597	12	AF529846	AF529846 Crictetu
37	592.2	21.0	597	12	AF529847	AF529847 Crictetu
38	592.2	21.0	597	12	AF529848	AF529848 Crictetu
39	592.2	21.0	597	12	AF529849	AF529849 Crictetu
40	592.2	21.0	597	12	AF529850	AF529850 Crictetu
41	592.2	21.0	597	12	AF529851	AF529851 Crictetu
42	592.2	21.0	597	12	AF529852	AF529852 Crictetu
43	592.2	21.0	597	12	AF529853	AF529853 Crictetu
44	592.2	21.0	597	12	AF529854	AF529854 Crictetu
45	586.2	20.8	591	12	AF529855	AF529855 Crictetu

ALIGNMENTS

RESULT 1  
AB040431  
LOCUS  
DEFINITION Homo sapiens AID mRNA for activation-induced cytidine deaminase,  
complete CDS.  
ACCESSION AB040431  
VERSION AB040431.1 GI:9988409  
KEYWORDS AID; activation-induced cytidine deaminase; Human AID.  
SOURCE Homo sapiens CDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (sites)  
AUTHORS Muto,T., Muramatsu,M., Taniwaki,M., Kinoshita,K. and Honjo,T.  
TITLE Isolation, tissue distribution, and chromosomal localization of the

JOURNAL  
MEDLINE  
REFERENCE  
AUTHORS

human activation-induced cytidine deaminase (AID) gene  
Genomics 68 (1), 85-88 (2000)

2 (sites)

Revy, P., Muto, T., Levv, Y., Geissmann, F., Plebani, A., Sanal, O.,  
Catalan, N., Forveille, M., Dufourcq-Lageouse, K., Gennery, A.,  
Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A. G., Brousse, N.,  
Muramatsu, M., Notarangelo, L. D., Kinoshita, K., Honjo, T., Fischer, A.  
and Durandy, A.

Activation-induced cytidine deaminase (AID) deficiency causes the  
autosomal recessive form of the Hyper-IgM syndrome (HIM2)

Cell 102 (5), 565-575 (2000)

JOURNAL  
MEDLINE  
REFERENCE

3 (bases 1 to 2791)

Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.  
Direct Submission  
Submitted (18-NAR-2000) Tasuku Honjo, Kyoto University, Department  
of Medical Chemistry, Faculty of Medicine; Yoshida, Sakyo-ku,  
Kyoto, Kyoto 606-8501, Japan (E-mail: honjo@four.med.kyoto-u.ac.jp,  
Tel: 81-75-753-4371 (ex. 4371), Fax: 81-75-753-4388)

Location/Qualifiers

1. .2791

/organism="Homo sapiens"

/db\_xref="taxon:9606"

1. .2791

/gene="AID"

77. .673

/gene="AID"

/codon\_start=1

/product="activation-induced cytidine deaminase"

/protein\_id="BAB12721.1"

/db\_xref="GI:9988410"

/translation="MSLLMNRKLYQFNVRWAKRRRTYLCYVVKRRDSATFSL  
DFGLRNKNGCHVELFLRYISDMIDPGRCYRVWTFKSPCYDCARHVPDLRNP  
NLSRIETARLYCEDRKAPEGLRLHRAGVQIAIMTKDYFYCNWTFVENHERTFK  
AWGLHNSVRLSRKRLPLDYEDVDLDAFRTLGL"

BASE COUNT 842 a 548 c 625 g 776 t

ORIGIN

Query Match 99.0%; Score 2791; DB 9; Length 2791;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2791; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 GAACATCATTAATGAAGTGAGATTTTCTGGCCCTGAGACTTGCAGGGAGCAAGA 63  
Dy 1 GAACATCATTAATGAAGTGAGATTTTCTGGCCCTGAGACTTGCAGGGAGCAAGA 60  
Qy 64 CACTCTGGACACCACATATGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATT 123  
Dy 61 CACTCTGGACACCACATATGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATT 120  
Qy 124 CAAAAATGTCGCTGGCTAAGGTCGCGCTGAGACCTACCTGCTGCTAGTGAAGAG 183  
Dy 121 CAAAAATGTCGCTGGCTAAGGTCGCGCTGAGACCTACCTGCTGCTAGTGAAGAG 180  
Qy 184 GCGTGACAGTGCTACATCCCTTTTCACTGGACTTTGGTTATPCITTCGCAATAAGAACGCGTG 243  
Dy 181 GCGTGACAGTGCTACATCCCTTTTCACTGGACTTTGGTTATPCITTCGCAATAAGAACGCGTG 240  
Qy 244 CCAGCTGGGAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTG 303  
Dy 241 CCAGCTGGGAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTG 300  
Qy 304 CTACCGCTGACCTGGTTACCTTCCTGGAGCCCTGCTAGCAGCTGTGCCGAGATGGC 363  
Dy 301 CTACCGCTGACCTGGTTACCTTCCTGGAGCCCTGCTAGCAGCTGTGCCGAGATGGC 360  
Qy 364 CGACTTTCTCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCCACCGCGGCTCTACTT 423  
Dy 361 CGACTTTCTCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCCACCGCGGCTCTACTT 420  
Qy 424 CTCTGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGGCGCTGCACCGCGCGGGTGCA 483

Db 421 CTGTGAGGACCGCAAGGCTGAGCCCGAGGGGCTGCGCGGCTGCAACCGCGGGGTGCA 480  
Qy 484 AATAGCCATCATGACCTTCAAGAGATTAATTTTACTGCTGGAATACTTTTGTAGAAAACCA 543  
Dy 481 AATAGCCATCATGACCTTCAAGAGATTAATTTTACTGCTGGAATACTTTTGTAGAAAACCA 540  
Qy 544 TGAAGAACCTTTCAAGGCTTGGGAAGGCTGCATGAAAATTCAGTTCTCTCTCCAGACA 603  
Dy 541 TGAAGAACCTTTCAAGGCTTGGGAAGGCTGCATGAAAATTCAGTTCTCTCTCCAGACA 600  
Qy 604 GCTTCGCGCATCCTTTTGGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTAC 663  
Dy 601 GCTTCGCGCATCCTTTTGGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTAC 660  
Qy 664 TTTGGGACTTTGATAGCAACTTCCAGGAATGTCCACACCATGAAATATCTCTGCTCAAG 723  
Dy 661 TTTGGGACTTTGATAGCAACTTCCAGGAATGTCCACACCATGAAATATCTCTGCTCAAG 720  
Qy 724 ACAGTGGATAAAAACAGTCCCTTCAAGTCTTCTCTGTTTATTTCTTCAACTCTCACTTT 783  
Dy 721 ACAGTGGATAAAAACAGTCCCTTCAAGTCTTCTCTGTTTATTTCTTCAACTCTCACTTT 780  
Qy 784 CTTAGAGTTTACAGAAAAATATTTATATAGGACTCTTTAAAAGATCTATGCTCTGAAA 843  
Dy 781 CTTAGAGTTTACAGAAAAATATTTATATAGGACTCTTTAAAAGATCTATGCTCTGAAA 840  
Qy 844 ATAGAGAAGCAACAGGCTGCGCAGGAGCTGCTGCAATTGGTGCAGTTTGAATGCA 903  
Dy 841 ATAGAGAAGCAACAGGCTGCGCAGGAGCTGCTGCAATTGGTGCAGTTTGAATGCA 900  
Qy 904 ACATTGTCCCTACTTGGGAATAACAGAACTGCAGGACCTGGGAGCATCCTAAAAGTGCAA 963  
Dy 901 ACATTGTCCCTACTTGGGAATAACAGAACTGCAGGACCTGGGAGCATCCTAAAAGTGCAA 960  
Qy 964 CGTTTTCTATGACTTTTAGTAGGATGAGAGCAAGGTAGATCTCTAAAAGCATGGTG 1023  
Dy 961 CGTTTTCTATGACTTTTAGTAGGATGAGAGCAAGGTAGATCTCTAAAAGCATGGTG 1020  
Qy 1024 AGAGGATCAAAATGTTTTATATCAACATCCTTTATTTGATTGATTGAGTTAAACAGT 1083  
Dy 1021 AGAGGATCAAAATGTTTTATATCAACATCCTTTATTTGATTGATTGAGTTAAACAGT 1080  
Qy 1084 GGTGTTAGTGATAGATTTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAACACAAA 1143  
Dy 1081 GGTGTTAGTGATAGATTTTCTATTCTTTTCCCTTGACGTTTACTTTCAAGTAACACAAA 1140  
Qy 1144 CTCTTCCATCAGGCCATGATCTATAGGACCTCTTAATGAGAGTATCTGGGTGATTGTGAC 1203  
Dy 1141 CTCTTCCATCAGGCCATGATCTATAGGACCTCTTAATGAGAGTATCTGGGTGATTGTGAC 1200  
Qy 1204 CCCAAACCATCTCTCCAAAGCATTAATATCCAATCATGCGCTGTATGTTTAAATCAGCAG 1263  
Dy 1201 CCCAAACCATCTCTCCAAAGCATTAATATCCAATCATGCGCTGTATGTTTAAATCAGCAG 1260  
Qy 1264 AAGCATGTTTTTATGTTGTGACAAAAGAGATTGTTATGGGTGGGATGAGGATATAGAC 1323  
Dy 1261 AAGCATGTTTTTATGTTGTGACAAAAGAGATTGTTATGGGTGGGATGAGGATATAGAC 1320  
Qy 1324 CATGCATGGTCACCTTCAAGCTACTTTAATAAAGGATCTTAAATGGGACAGGAGGACTGT 1383  
Dy 1321 CATGCATGGTCACCTTCAAGCTACTTTAATAAAGGATCTTAAATGGGACAGGAGGACTGT 1380  
Qy 1384 GAACAAGACACCCCTAATAATGTTGTTGATGCTCAAGTACGAAATCTTCTGGAACGCAAA 1443  
Dy 1381 GAACAAGACACCCCTAATAATGTTGTTGATGCTCAAGTACGAAATCTTCTGGAACGCAAA 1440  
Qy 1444 CTCCTTTAAGGAAGTCCCTAATAATTTAGAAACACCCACCAAACTTCACATATCATTAATAGCA 1503  
Dy 1441 CTCCTTTAAGGAAGTCCCTAATAATTTAGAAACACCCACCAAACTTCACATATCATTAATAGCA 1500  
Qy 1504 AACAAATGGGAAGGATGCTTCAATCTTGGGAGAGGAAATCTATGCTCTCGTGGG 1563  
Dy 1501 AACAAATGGGAAGGATGCTTCAATCTTGGGAGAGGAAATCTATTTGGCTCTCGTGGG 1560



Db	8950	CTACAGCCCTGATGAGGTTGATGACATTACGAGCGCATTCGTA	1757	GTTCAGGCCAGCCTGGGCAACATACAGATCCCTCTCTCAAAAAA	1816
Qy	677	TAGCAACTTCAGGAATGTCACACAGCATGAATATCTGCTGAAGACAGTGGATAAA	10090	GTTCAGGCCAGCCTGGGCAACATACAGATCCCTCTCTCAAAAAA	10149
Db	9010	TAGCAACTTCAGGAATGTCACACAGCATGAATATCTGCTGAAGACAGTGGATAAA	1817	AAGAAAGAGAGAGGCCCGGGCGTGGTGTCTCAGGCTGTAAATCCAGACACTTTGGAGGC	1876
Qy	737	AACAGTCTCTCAAGTCTCTGTTTATCTTCAACTCTCACATTTCTTAGAGTTTACA	10150	AAGAAAGAGAGAGGCCCGGGCGTGGTGTCTCAGGCTGTAAATCCAGACACTTTGGAGGC	10209
Db	9070	AACAGTCTCTCAAGTCTCTGTTTATCTTCAACTCTCACATTTCTTAGAGTTTACA	1877	CGAGCCGGCGGATCACCTGTGGTCAGGAGTTTGAGACAGCCTGGCCCAACATGCAAAA	1936
Qy	797	GAATAAATATTTATACGACTCTTTAAAAAGATCTATGTCTGAAAAATAGAGAAGAAC	10210	CGAGCCGGCGGATCACCTGTGGTCAGGAGTTTGAGACAGCCTGGCCCAACATGCAAAA	10269
Db	9130	GAATAAATATTTATACGACTCTTTAAAAAGATCTATGTCTGAAAAATAGAGAAGAAC	1937	CCCCCTCTGTACTCAAAAATTTAGCCAGGCTGTAGCAGGACCTGTAAATCCC	1996
Qy	857	ACAGTCTGGCCAGGAGCGTGTGCAATTTGGTGAGTTTGAATGCAACATTTGCCCTA	10270	CCCCCTCTGTACTCAAAAATTTAGCCAGGCTGTAGCAGGACCTGTAAATCCC	10329
Db	9190	ACAGTCTGGCCAGGAGCGTGTGCAATTTGGTGAGTTTGAATGCAACATTTGCCCTA	1997	AGCTACTTTGGGAGGCTGAGCGAGGAGATCGCTTGAACCCAGGAGTGGAGTTCACATA	2056
Qy	917	CTGGGAATACAGAACTGCAGGACCTGGGAGCATCTTAAGTGTCAAGTCTTCTATGA	10330	AGCTACTTTGGGAGGCTGAGCGAGGAGATCGCTTGAACCCAGGAGTGGAGTTCACATA	20389
Db	9250	CTGGGAATACAGAACTGCAGGACCTGGGAGCATCTTAAGTGTCAAGTCTTCTATGA	2057	AGCTGAGATCGTCCGTTGCACTCCAGCTTGGGCGACAAGACAGACTCTGTCTCAGAA	2116
Qy	977	CTTTTAGGTAGGATGAGACAGAGGTAGATCCCTAAAAAGCATGGTGAGAGATCAAAATG	10390	AGCTGAGATCGTCCGTTGCACTCCAGCTTGGGCGACAAGACAGACTCTGTCTCAGAA	10449
Db	9310	CTTTTAGGTAGGATGAGACAGAGGTAGATCCCTAAAAAGCATGGTGAGAGATCAAAATG	2117	AAAAAATAAAG	2176
Qy	1037	TTTTTATATCAACATCTTTATTTATTTGATTTCAATTTGAGTTTAAACAGTGGTGTAGTGATA	10450	AAAAAATAAAG	2176
Db	9370	TTTTTATATCAACATCTTTATTTATTTGATTTCAATTTGAGTTTAAACAGTGGTGTAGTGATA	2177	AAGCATTTGCAAGGAAATTTGCTTTATCCAAACAAATTTAAGGAGCCAAATTAAGGATCCC	2236
Qy	1097	GATTTTCTATTTCTTTCCCTTACGTTTCAAGTATCAAGTATCAACAAACTCTTCCATCAGG	10510	AAGCATTTGCAAGGAAATTTGCTTTATCCAAACAAATTTAAGGAGCCAAATTAAGGATCCC	10569
Db	9430	GATTTTCTATTTCTTTCCCTTACGTTTCAAGTATCAAGTATCAACAAACTCTTCCATCAGG	2237	TATTTGTCTCTTTTGGTGCTTATTTGTCCTTAACTGCTTTTGCAGAGTGAAGAAATA	2296
Qy	1157	CCATGATCTATAGGACCTCCTTAATGAGAGTATCTGGGTGATTTGACCCCAAAACCATCTC	10570	TATTTGTCTCTTTTGGTGCTTATTTGTCCTTAACTGCTTTTGCAGAGTGAAGAAATA	10629
Db	9490	CCATGATCTATAGGACCTCCTTAATGAGAGTATCTGGGTGATTTGACCCCAAAACCATCTC	2297	TTCAGATAACCATPATCCCTGTGCCGTTATTACCTAGCAACCCCTTGAAGTGAAGATGAGC	2356
Qy	1217	TCCAAAGCATTAATATCCATCATGCGCTGTATGTTTAAATCAGCAGAGCATGTTTATA	10630	TTCAGATAACCATPATCCCTGTGCCGTTATTACCTAGCAACCCCTTGAAGTGAAGATGAGC	10689
Db	9550	TCCAAAGCATTAATATCCATCATGCGCTGTATGTTTAAATCAGCAGAGCATGTTTATA	2357	AGATCCACAGGAAACCTTGAATGCACAACTGTCTTATTTAACTCTTATGCTACATAAGTT	2416
Qy	1277	TGTTTGTACAAAAGAGATGTTTATGGGTGGGATGGAGGTATAGACCATGCAATGGTCAC	10690	AGATCCACAGGAAACCTTGAATGCACAACTGTCTTATTTAACTCTTATGCTACATAAGTT	10749
Db	9610	TGTTTGTACAAAAGAGATGTTTATGGGTGGGATGGAGGTATAGACCATGCAATGGTCAC	2417	TGTAAGAGAGTTAAAAATTTGTTACTTCATGTATTTCATTTATTTATTTATTTTTCGCT	2476
Qy	1337	CTTCAAGCTACTTTAATAAGGATCTTAAATGGCGAGGAGGACTGTGAACAAGACACCC	10750	TGTAAGAGAGTTAAAAATTTGTTACTTCATGTATTTCATTTATTTATTTATTTTTCGCT	10809
Db	9670	CTTCAAGCTACTTTAATAAGGATCTTAAATGGCGAGGAGGACTGTGAACAAGACACCC	2477	CTAATGATTTTTTATTAAACATGATTTCCCTTTCTGTATATATTGAAATGGAGTCTCAAGC	2536
Qy	1397	TAATAATGGGTGATGCTCGAAGTACCAATCTCTGGAACGCAAACTCTTTAAGGAA	10810	CTAATGATTTTTTATTAAACATGATTTCCCTTTCTGTATATATTGAAATGGAGTCTCAAGC	10869
Db	9730	TAATAATGGGTGATGCTCGAAGTACCAATCTCTGGAACGCAAACTCTTTAAGGAA	2537	TTCATAAATTTATAACCTTTTGAAGTATGTTCTTAATAACAAGTATGTAATTTGTAACATTC	2596
Qy	1457	GTCCTAAATTTAGAAACACCCACAACTTCACATATCATAAATAGCAAACTTTGGAAG	10870	TTCATAAATTTATAACCTTTTGAAGTATGTTCTTAATAACAAGTATGTAATTTGTAACATTC	10929
Db	9790	GTCCTAAATTTAGAAACACCCACAACTTCACATATCATAAATAGCAAACTTTGGAAG	2597	AGTAATGGTCTACGAGCCATTTCTCTGATTTTTAGTAACTTTTATGACAGCAAAAT	2656
Qy	1517	AAGTGTCTGAATTTGGGAGAGGAAATCTATTTGGCTCTCGTGGTCTCTTTCATCTCA	10930	AGTAATGGTCTACGAGCCATTTCTCTGATTTTTAGTAACTTTTATGACAGCAAAAT	10989
Db	9850	AAGTGTCTGAATTTGGGAGAGGAAATCTATTTGGCTCTCGTGGTCTCTTTCATCTCA	2657	TGCTTCTGGCTCACTTTCAATAGTAAATGATAAATAAATAAATTTTGGAGCTGTGAAG	2716
Qy	1577	GAATGCAATCAGGTCAAGTGTGCTACATTTTGTATGTTGATGCTTCTCCCAAG	10990	TGCTTCTGGCTCACTTTCAATAGTAAATGATAAATAAATAAATTTTGGAGCTGTGAAG	11049
Db	9910	GAATGCAATCAGGTCAAGTGTGCTACATTTTGTATGTTGATGCTTCTCCCAAG	2717	ATAAAAATACCAATAAATAAATAAAGTGTATTAATCAAGCTTAAATTAATAAATCAG	2776
Qy	1637	GTATATTAATATATAAGAGTTGTGACAAAAAGAAATGATAAAGCTGCGAACCTGGC	11050	ATAAAAATACCAATAAATAAATAAAGTGTATTAATCAAGCTTAAATTAATAAATCAG	2776
Db	9970	GTATATTAATATATAAGAGTTGTGACAAAAAGAAATGATAAAGCTGCGAACCTGGC	2777	TATGATGGAATAAACTTGAAA	2797
Qy	1697	ACAGGCTCATAGTTCTAGTGTCTGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGT	11110	TATGATGGAATAAACTTGAGA	11130
Db	10030	ACAGGCTCATAGTTCTAGTGTCTGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGT			





repeat_region	3296..3328	/rpt_family=" (TTTC)n"	Qy	797	GAATAATATTTATATACGACTCTTTAAAAAGATCTATGCTTTGAAATAGAGAAGAAC	856
repeat_region	4097..4249	/rpt_family="AluSq"	Db	43673	GAATAATATTTATATACGACTCTTTAAAAAGATCTATGCTTTGAAATAGAGAAGAAC	43732
repeat_region	4360..4654	/rpt_family="AluY"	Qy	857	ACAGGTCTGGCCAGGGACGTGCTGCAATTTGGTGAGATTTTGAATGCAACATTTGCCCTTA	916
repeat_region	complement(5140..5262)	/rpt_family="FLAM_C"	Db	43733	ACAGGTCTGGCCAGGGACGTGCTGCAATTTGGTGAGATTTTGAATGCAACATTTGCCCTTA	43792
repeat_region	5275..5304	/rpt_family="AT_rich"	Qy	917	CTGGGAATAACAGAACTGCAGGACCTGGAGACCTCTAAAGTGTCAACGTTTTTCTATGA	976
repeat_region	5305..5563	/rpt_family="AluJo"	Db	43793	CTGGGAATAACAGAACTGCAGGACCTGGAGACCTCTAAAGTGTCAACGTTTTTCTATGA	43852
repeat_region	5564..5599	/rpt_family="AT_rich"	Qy	977	CTTTTAGGTAGGTAGGATGAGAGAGAGTAGATCTCTAAAAAGCATGGTGAGAGATCAAAATG	1036
repeat_region	5644..5733	/rpt_family="MSTD"	Db	43853	CTTTTAGGTAGGTAGGATGAGAGAGAGTAGATCTCTAAAAAGCATGGTGAGAGATCAAAATG	43912
repeat_region	5734..6012	/rpt_family="AluSq"	Qy	1037	TTTTTATATCAACATCCTTTATTTATTTGATTCATTTGAGTTTACAGTTGGTGTAGTATATA	1096
repeat_region	6020..6336	/rpt_family="AluSx"	Db	43913	TTTTTATATCAACATCCTTTATTTATTTGATTCATTTGAGTTTACAGTTGGTGTAGTATATA	43972
repeat_region	6337..6629	/rpt_family="MSTD"	Qy	1097	GATTTTCTATTTCTTTCCCTTTGACGTTTACTTTCAAGTAAACACAAACTCTTCCATCAGG	1156
repeat_region	6630..6650	/rpt_family="(CAA)n"	Db	43973	GATTTTCTATTTCTTTCCCTTTGACGTTTACTTTCAAGTAAACACAAACTCTTCCATCAGG	44032
repeat_region	complement(6711..6817)	/rpt_family="MIR"	Qy	1157	CCATGATCTATAGAACCTCTCTTAATGAGAGTATCTGGGTGATGTGACCCCAACCATCTCT	1216
repeat_region	complement(7033..7327)	/rpt_family="AluSx"	Db	44033	CCATGATCTATAGAACCTCTCTTAATGAGAGTATCTGGGTGATGTGACCCCAACCATCTCT	44092
repeat_region	7607..7745	/rpt_family="FLAM_A"	Qy	1217	TCCAAGCATTAATCAACATCATGCGCTGTATCTTTTAAACACGAGAGCATCTTTTATA	1276
repeat_region	7980..8047	/rpt_family="MIR"	Db	44093	TCCAAGCATTAATCAACATCATGCGCTGTATCTTTTAAACACGAGAGCATCTTTTATA	44152
repeat_region	complement(8050..8549)	/rpt_family="LTR47A"	Qy	1277	TGTTTGTACAAAAGAAGATTCTTATGGTGGGGATGGAGGTATAGACCATGCATGGTCAC	1336
repeat_region	8567..8698	/rpt_family="MIR"	Db	44153	TGTTTGTACAAAAGAAGATTCTTATGGTGGGGATGGAGGTATAGACCATGCATGGTCAC	44212
repeat_region	complement(8815..9101)	/rpt_family="AluSg"	Qy	1337	CTTCAAGCTACTTTAATAAAGGATCTTAAATGGCAGGAGGACTGTGAACAACACACACC	1396
repeat_region	complement(11227..12175)	/rpt_family="LTR5"	Db	44213	CTTCAAGCTACTTTAATAAAGGATCTTAAATGGCAGGAGGACTGTGAACAACACACACC	44272
STS	12625..12900	/standard_name="87432"	Qy	1397	TAATAAGTGGTGTGATGCTGAAGTAGCAAAATCTTTCTGAAACGCAAACTCTTTTAAGGAA	1456
repeat_region	complement(13547..13664)	/rpt_family="LIMB5"	Db	44273	TAATAAGTGGTGTGATGCTGAAGTAGCAAAATCTTTCTGAAACGCAAACTCTTTTAAGGAA	44332
repeat_region	complement(13665..13979)	/rpt_family="AluY"	Qy	1457	GTCCCTAATTTAGAAAACCCACACAACTTCACATATCATATATTAAGCAAACTTGAAGG	1516
repeat_region	complement(13980..14193)	/rpt_family="LIMB5"	Db	44333	GTCCCTAATTTAGAAAACCCACACAACTTCACATATCATATATTAAGCAAACTTGAAGG	44392
repeat_region	14622..14924	/rpt_family="AluY"	Qy	1517	AAGTTGCTTGAATGTTGGGGAGAGAAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA	1576
repeat_region	complement(15813..17136)	/rpt_family="LIP48"	Db	44393	AAGTTGCTTGAATGTTGGGGAGAGAAAATCTATTGGCTCTCGTGGGTCTCTTCATCTCA	44452
repeat_region	complement(17272..17395)	/rpt_family="FLAM_C"	Qy	1577	GAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGTCTCTCCCAAG	1636
repeat_region	17426..17575	/rpt_family="MIR"	Db	44453	GAATGCCAATCAGGTCAAGGTTTGCTACATTTTGTATGTGTGTGTCTCTCCCAAG	44512
repeat_region	19021..19314	/rpt_family="MIR"	Qy	1637	GTATATTAACATATATAAGAGAGTTGTGACAAAACAGAAATGATAAGCTGCGAACCGTGGC	1696
Query Match	77.2%; Score 2174.6; DB 9; Length 71132;		Db	44513	GTATATTAACATATATAAGAGAGTTGTGACAAAACAGAAATGATAAGCTGCGAACCGTGGC	44572
Best Local Similarity 99.8%; Pred. No. 0;			Qy	1697	ACAGCTCATAGTTCTAGCTGCTGGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGT	1756
Matches 2177; Conservative 0; Mismatches 4; Indels 0; Gaps 0;			Db	44573	ACAGCTCATAGTTCTAGCTGCTGGGAGGTTGAGGAGGAGGATGGCTTGAACACAGGT	44632
Qy	617	CTTTTGGCCCTGTATGAGTTGATGACTTACGAGACGATTTCCGTACTTTGGGACTTTGA	Qy	1757	GTTCAAGCCAGCCTGGGCAACATAAACAGATCTGTCTCTCAAAAAAATAAAAAA	1816
Db	43493	CTACAGCCCTGTATGAGTTGATGACTTACGAGACGATTTCCGTACTTTGGGACTTTGA	Db	44633	GTTCAAGCCAGCCTGGGCAACATAAACAGATCTGTCTCTCAAAAAAATAAAAAA	44692
Qy	677	TAGCAACTTCAGGATGTCACACGATGAATATCTCTGCTGAAGACAGTGGATAAA	Qy	1817	AAGAAAGAGAGAGGCGCGGCGTGTGCTCACGCCCTTAATCCAGCACATTTGGGAGGC	1876
Db	43553	TAGCAACTTCAGGATGTCACACGATGAATATCTCTGCTGAAGACAGTGGATAAA	Db	44693	AAGAAAGAGAGAGGCGCGGCGTGTGCTCACGCCCTTAATCCAGCACATTTGGGAGGC	44752

```
QY 1877 CGAGCCGGCGGATCACCTGTGGTCAGGAGTTTGAGACACAGCCTGGCCAACTGCGCAAA 1936
Db 44753 CGAGCCGGCGGATCACCTGTGGTCAGGAGTTTGAGACACAGCCTGGCCAACTGCGCAAA 44812
QY 1937 CCCCCTGTACTCAAAATGCAAAATTTAGCCAGCGGTGGTAGCAGGACACCTGTATCC 1996
Db 44813 CCCCCTGTACTCAAAATGCAAAATTTAGCCAGCGGTGGTAGCAGGACACCTGTATCC 44872
QY 1997 AGCTACTTGGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGTTCAGTA 2056
Db 44873 AGCTACTTGGGAGGCTGAGGAGGAGAAATCGCTTGAACCCAGGAGGTGGAGGTTCAGTA 44932
QY 2057 AGCTGAGATCGGCCGTTGCACTCCAGCCTGGCGACAGCAAGCAAGCACTCTCTCAGAA 2116
Db 44933 AGCTGAGATCGGCCGTTGCACTCCAGCCTGGCGACAGCAAGCAAGCACTCTCTCAGAA 44992
QY 2117 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2176
Db 44993 AAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 45052
QY 2177 AAGCATTCGAAGGAAATGTGCTTTATCCACAAATGTAAGGAGCCAAATAGGGATGCC 2236
Db 45053 AAGCATTCGAAGGAAATGTGCTTTATCCACAAATGTAAGGAGCCAAATAGGGATGCC 45112
QY 2237 TATTTGTCTCTTTTGGTGCTATTGTGTCCTAACCACTGCTTTGACAGTGAGAAAAATA 2296
Db 45113 TATTTGTCTCTTTTGGTGCTATTGTGTCCTAACCACTGCTTTGACAGTGAGAAAAATA 45172
QY 2297 TTCAGAAATCAATATCCCTGCGGTATTACCTAGCAACCCCTGCAATGAAGATGAGC 2356
Db 45173 TTCAGAAATCAATATCCCTGCGGTATTACCTAGCAACCCCTGCAATGAAGATGAGC 45232
QY 2357 AGATCCACAGGAAACTTGAATGCACAACTGCTTATTTAATCTTATTTGATACATAAGTT 2416
Db 45233 AGATCCACAGGAAACTTGAATGCACAACTGCTTATTTAATCTTATTTGATACATAAGTT 45292
QY 2417 TGTAAAGAGTTAAAAATTTGTTACTTCATGTATTCATTTATATTTATATTTTGGCT 2476
Db 45293 TGTAAAGAGTTAAAAATTTGTTACTTCATGTATTCATTTATATTTATTTTGGCT 45352
QY 2477 CTAATGATTTTTTATACATGATTTCCCTTTTCTGATATTTGAAATGAGTCTCAAGC 2536
Db 45353 CTAATGATTTTTTATACATGATTTCCCTTTTCTGATATTTGAAATGAGTCTCAAGC 45412
QY 2537 TTCATAAATTTAATCTTTAGAAATGATTCATTAACAACGATGTAATTTGAATTCG 2596
Db 45413 TTCATAAATTTAATCTTTAGAAATGATTCATTAACAACGATGTAATTTGAATTCG 45472
QY 2597 AGTAATGGTGCTACGAAGCCATTTCTCTGATTTTTAGTAACTTTTATGACAGCAAT 2656
Db 45473 AGTAATGGTGCTACGAAGCCATTTCTCTGATTTTTAGTAACTTTTATGACAGCAAT 45532
QY 2657 TCGTCTTGCTCACITTCATCAGTTAAATGAATGAATAAATTTTGAAGCTGTGAAG 2716
Db 45533 TCGTCTTGCTCACITTCATCAGTTAAATGAATGAATAAATTTTGAAGCTGTGAAG 45592
QY 2717 ATAAATACCAATAAATTAATATAAGTGTATTAATGAAGTTAAATATAAATATCAG 2776
Db 45593 ATAAATACCAATAAATTAATATAAAGTGTATTAATGAAGTTAAATATAAATATCAG 45652
QY 2777 TATGATGGAATAAATTTGAAA 2797
Db 45653 TATGATGGAATAAATTTGAGA 45673

RESULT 4
BC006296 1837 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, activation-induced cytidine deaminase, clone
DEFINITION MGC:12911 IMAGE:4054915, mRNA, complete cds.
ACCESSION BC006296
VERSION BC006296.1 GI:13623400
KEYWORDS MGC.
```

```
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1837)
AUTHORS Strausberg,R.
TITLE Direct Submission
JOURNAL Submitted (09-APR-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT Contact: MGC help desk
Email: cgabp@remail.nih.gov
Tissue procurement: Louis Staudt
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
nisc_mgc@nhgri.nih.gov
Sherchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R.,
Lim,M., Maduro,Q.L., Masiello,C., Mastrian,S.D., McCloskey,J.C.,
McDowell,J., Pearson,R., Snyder,B., Stantripop,S., Thomas,P.J.,
Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
Zhang,L.-H. and Green,E.D.
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 17 Row: a Column: 1
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 9988409.
FEATURES
Location/Qualifiers
1..1837
/organism="Homo sapiens"
/db_xref="LocusID:57379"
/db_xref="taxon:9606"
/clone="MGC:12911 IMAGE:4054915"
/tissue_type="Primary B-cells from Tonsils"
/gene_lib="NIH_MGC_48"
/lab_host="DH10B-R"
/note="vector: pOTB7"
86..682
/codon_start=1
/product="activation-induced cytidine deaminase"
/protein_id="AAH06296.1"
/db_xref="GI:13623401"
/translation="MDSLMMNRKFLYQFNKRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFLAYISDWLDLPGRCYRVWTFWSKPCYDCARHVDPLRGNP
NLSLRIFTARLYFCEDRKAEPEGLRLRHAGVQIALMTPEKDFYCNWTFVENHRTFK
AWEGHRENSVRSLRQLRILLPLYEVDLDRAPFTLGL"
BASE COUNT 530 a 387 c 421 g 499 t
ORIGIN
Query Match 64.6%; Score 1820; DB 9; Length 1837;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1823; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 AGAGAACCACATTAATTGAAGTGAGATTTTCTGGCCTGAGACTTTCGAGGGAGGCAAGA 60
Db 7 AGGGAACCACATTAATTGAAGTGAGATTTTCTGGCCTGAGACTTTCGAGGGAGGCAAGA 66
QY 61 AGACACTCTGGACACCACTATGACACGCTCTTGTATGAACCGGAGGAAGTTTCTTTACCA 120
Db 67 AGACACTCTGGACACCACTATGACACGCTCTTGTATGAACCGGAGGAAGTTTCTTTACCA 126
QY 121 ATTCAAAATGTCGGCTGGCTAAGGTTCGGCTGAGACCTACCTGCTGCTACCTAGTAA 180
Db 127 ATTCAAAAATGTCGGCTGGCTAAGGTTCGGCTGAGACCTACCTGCTGCTACCTAGTAA 186
```

Qy	181	GAGCGTGACAGTCTACATCCTTTTCACCTGGACTTTGGTATCTTCCCAATAAGACGG	240
Db	187	GAGCGTGACAGTCTACATCCTTTTCACCTGGACTTTGGTATCTTCCCAATAAGACGG	246
Qy	241	CTGCGAGTGGAAATGCTCTTCCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCG	300
Db	247	CTGCCAGTGGAAATGCTCTTCCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCCG	306
Qy	301	CTGCTACCGCGTCACTGGTTCACTCTCCCTGAGGCCCTGCTAGGACTGTGCCGACATGT	360
Db	307	CTGCTACCGCGTCACTGGTTCACTCTCCCTGAGGCCCTGCTAGGACTGTGCCGACATGT	366
Qy	361	GGCGGACTTTCTGGGAGGGAACCCCAACCTCAGTCTGAGGACTTTCACCGGGCGCTCTA	420
Db	367	GGCGGACTTTCTGGGAGGGAACCCCAACCTCAGTCTGAGGACTTTCACCGGGCGCTCTA	426
Qy	421	CTTCTGTGAGACCGCAAGGCTGAGCCCGAGGGGCTGCGGGGCTGCACCGCGCGGGT	480
Db	427	CTTCTGTGAGACCGCAAGGCTGAGCCCGAGGGGCTGCGGGGCTGCACCGCGCGGGT	486
Qy	481	GCAAAATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAA	540
Db	487	GCAAAATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAA	546
Qy	541	CCATGAAAGAACTTTCAAGCCCTGGGAGGCTGCATGAAATTCAGTTCGCTCTCCAG	600
Db	547	CCATGAAAGAACTTTCAAGCCCTGGGAGGCTGCATGAAATTCAGTTCGCTCTCCAG	606
Qy	601	ACAGCTTCGGCGCATCTTTTCCCTGATAGGTTGATGACTTACGAGACGCATTTCCG	660
Db	607	ACAGCTTCGGCGCATCTTTTCCCTGATAGGTTGATGACTTACGAGACGCATTTCCG	666
Qy	661	TACTTTGGGACTTTGATAGCAACTTCCAGGAATGTACACACGATGAAATATCTCTGCTG	720
Db	667	TACTTTGGGACTTTGATAGCAACTTCCAGGAATGTACACACGATGAAATATCTCTGCTG	726
Qy	721	AAGACAGTGATAAAAACAGTCTTCAAGTCTCTCTGTTTATTTATTTCTTCACTCTCAC	780
Db	727	AAGACAGTGATAAAAACAGTCTTCAAGTCTCTCTGTTTATTTATTTCTTCACTCTCAC	786
Qy	781	TTTCTTAGCTTTACAGAAAAATATTTATATAGGACTTTTAAAAAGATCTATGCTCTG	840
Db	787	TTTCTTAGCTTTACAGAAAAATATTTATATAGGACTTTTAAAAAGATCTATGCTCTG	846
Qy	841	AAAAATAGAGGAAGAACACAGTCTGCGCAGGAGCGTCTGCAATTTGGTGCAAGTTTGAAT	900
Db	847	AAAAATAGAGGAAGAACACAGTCTGCGCAGGAGCGTCTGCAATTTGGTGCAAGTTTGAAT	906
Qy	901	GCAACATTTGCCCTACTGCGAATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGT	960
Db	907	GCAACATTTGCCCTACTGCGAATAACAGAACTGCAGGACCTGGGAGCATCCTAAAGTGT	966
Qy	961	CAACGTTTTTCTATGACTTTTAGTGTAGGATGAGAGCAGAGGTAGATCCTAAAAAGCATG	1020
Db	967	CAACGTTTTTCTATGACTTTTAGTGTAGGATGAGAGCAGAGGTAGATCCTAAAAAGCATG	1026
Qy	1021	GTGAGAGGATCAAAATGTTTTATATCAACATCCTTTATTTATTTCAATTTGAGTTAAC	1080
Db	1027	GTGAGAGGATCAAAATGTTTTATATCAACATCCTTTATTTATTTCAATTTGAGTTAAC	1086
Qy	1081	AGTGGTGTAGTATAGATTTTCTATTTCTTTTCCCTTGACGTTTACTTTCAAGTAACAC	1140
Db	1087	AGTGGTGTAGTATAGATTTTCTATTTCTTTTCCCTTGACGTTTACTTTCAAGTAACAC	1146
Qy	1141	AAACTCTTCCATCAGGCGCATGATCTATAGGACCTCCTAAATGAGAGTATCTGGGTGATTGT	1200
Db	1147	AAACTCTTCCATCAGGCGCATGATCTATAGGACCTCCTAAATGAGAGTATCTGGGTGATTGT	1206
Qy	1201	GACCCCAAAACCATCTCTCCAAAGCAATTAATCAATCATCGCTGTATCTTTTAATCAG	1260
Db	1207	GACCCCAAAACCATCTCTCCAAAGCAATTAATCAATCATCGCTGTATCTTTTAATCAG	1266
Qy	1261	CAGAAGCATGTTTTTATGTTTGTACAAAGAAAGATTGTTATGGTGGGGATGGAGGTATA	1320
Db	1267	CAGAAGCATGTTTTTATGTTTGTACAAAGAAAGATTGTTATGGTGGGGATGGAGGTATA	1326
Qy	1321	GACCATGCATGGTCCACCTTCAAGCTACTTTAATAAAGATCTTAAAAATGGCAGGAGGAC	1380
Db	1327	GACCATGCATGGTCCACCTTCAAGCTACTTTAATAAAGATCTTAAAAATGGCAGGAGGAC	1386
Qy	1381	TGTGAACAAGACACCCCTAAATAATGGTGTGATCTGAAGTAGCAAAATCTTCTGGAACGC	1440
Db	1387	TGTGAACAAGACACCCCTAAATAATGGTGTGATCTGAAGTAGCAAAATCTTCTGGAACGC	1446
Qy	1441	AAACTCTTTTAAAGAAAGTCCCTTAATTTAGAAACACCCACAAACCTTCACATATCATATTA	1500
Db	1447	AAACTCTTTTAAAGAAAGTCCCTTAATTTAGAAACACCCACAAACCTTCACATATCATATTA	1506
Qy	1501	GCAAAACAATTTGGAAGGAAGTTGCTTGAATGTTGGGAGAGGAAAAATCTATTGGCTCTCGT	1560
Db	1507	GCAAAACAATTTGGAAGGAAGTTGCTTGAATGTTGGGAGAGGAAAAATCTATTGGCTCTCGT	1566
Qy	1561	GGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGTCTACATTTTGTATGTGTGT	1620
Db	1567	GGGTCTCTTCATCTCAGAAATGCCAATCAGGTCAAGGTTTGTCTACATTTTGTATGTGTGT	1626
Qy	1621	GATGCTCTTCTCCAAAGGTATATTAACTATATAAGAGAGTTGTGACAAACAGAAATGATAA	1680
Db	1627	GATGCTCTTCTCCAAAGGTATATTAACTATATAAGAGAGTTGTGACAAACAGAAATGATAA	1686
Qy	1681	AGCTGCCAACCTGGCACAGCTCATAGTTCTAGCTGCTTGGGAGGTTTGAGGAGGGAGGA	1740
Db	1687	AGCTGCCAACCTGGCACAGCTCATAGTTCTAGCTGCTTGGGAGGTTTGAGGAGGGAGGA	1746
Qy	1741	TGCGTTGAACACAGAGGTGTTCAAGGCCAGCCTGGGCAACATAACAAAGATCTCTCTCAA	1800
Db	1747	TGCGTTGAACACAGAGGTGTTCAAGGCCAGCCTGGGCAACATAACAAAGATCTCTCTCAA	1806
Qy	1801	AAAAAAAAAAAAAAAAAAAAAGAGAGAGA	1828
Db	1807	AAAAAAAAAAAAAAAAAAAAA	1834
RESULT	5		
AF132979			
LOCUS	AF132979	2440 bp	linear
DEFINITION	Mus musculus activation-induced cytidine deaminase (AID) mRNA, complete cds.		
ACCESSION	AF132979		
VERSION	AF132979.1	GI:5305727	
KEYWORDS	Mus musculus.		
SOURCE	Mus musculus		
ORGANISM	Mus musculus		
REFERENCE	1 (bases 1 to 2440)		
AUTHORS	Muramatsu, M., Sankaranand, V.S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N.O. and Honjo, T.		
TITLE	Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells		
JOURNAL	J. Biol. Chem. 274 (26), 18470-18476 (1999)		
MEDLINE	9303612		
PUBMED	10373455		
REFERENCE	2 (bases 1 to 2440)		
AUTHORS	Muramatsu, M. and Honjo, T.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-NAR-1999) Medical Chemistry, Kyoto University, Konoe Yoshida Sakyo-ku, Kyoto 606-8501, Japan		
FEATURES	Location/Qualifiers		
source	1..2440		
gene	/organism="Mus musculus"		
	/db_xref="taxon:10090"		
	1..2440		
	/gene="AID"		
CDS	93..689		



source	1. .597 /organism="Homo sapiens" /db_xref="taxon:9606" 1. .597 /gene="AID" 1. .597 /gene="AID" /codon_start=1 /product="AID" /protein_id="AAM95415.1" /db_xref="GI:22297244" /translation="MDSLLMRKFLYQFNWAKGRRETYLCYVVKRDSATSFSL DFGYLRNKNCHVELLFLRYISDWDLDPGRCYRVWFTSWSPCYDCARHVADFLRGNP NLSLRIFETARLYFCEDRKAEPEGLRLRHAGVQIAIMTFKDYFCYWNTEFVENHERTEFK AWGLHENSVRLSRQLRRILLPLYEVDLRDAFTLGL"	129 a 164 c 154 g 150 t
BASE COUNT		
ORIGIN		
Query Match 21.1%; Score 595.4; DB 12; Length 597; Best Local Similarity 99.8%; Pred. No. 4.1e-107; Matches 596; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	80	ATGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGG 139
Db	1	ATGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGG 60
QY	140	GCTAAGGGTCGGCGTGAGACCTACCTGTGTCTAGCTAGTGAAGAGCGGTGACAGTGTCTACA 199
Db	61	GCTAAGGGTCGGCGTGAGACCTACCTGTGTCTAGCTAGTGAAGAGCGGTGACAGTGTCTACA 120
QY	200	TCCTTTTCACTGGACCTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 259
Db	121	TCCTTTTCACTGGACCTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 180
QY	260	TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGTCTACCGGTCACCTGG 319
Db	181	TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGTCTACCGGTCACCTGG 240
QY	320	TTCAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAAAGAACTTTCAAA 559
Db	421	TTCAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAAAGAACTTTCAAA 480
QY	560	GCTGGGAAGGGCTGCATGAAATTCAGTTCCAGACAGCTTCGACTTTGGGACTTTGA 676
Db	541	TTGCCCTCTATGAGGTTGATGACTTACGAGACGCAATTTTCGTACTTTGGGACTTTGA 597
RESULT 7		
AF529815		
LOCUS	596 bp	mRNA
DEFINITION	Homo sapiens clone Ramos 1 AID (AID)	linear PRI 19-AUG-2002
ACCESSION	AF529815	partial cds.
VERSION	AF529815.1	GI:22297217
KEYWORDS	human.	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	

REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 596) Martin, A. and Scharff, M.D. Somatic hypermutation of the AID transgene in B cells and non-B cells Unpublished 2 (bases 1 to 596) Martin, A. and Scharff, M.D. Direct Submission Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA Location/Qualifiers 1. .596 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="Ramos 1" 1. .>596 /gene="AID" 1. .>596 /gene="AID" /note="integrated into Burkitt's lymphoma cell line Ramos"	
FEATURES	source	
gene		
CDS		
BASE COUNT 128 a 163 c 155 g 150 t		
ORIGIN		
Query Match 21.1%; Score 594.4; DB 9; Length 596; Best Local Similarity 99.8%; Pred. No. 6.5e-107; Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		
QY	80	ATGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGG 139
Db	1	ATGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGG 60
QY	140	GCTAAGGGTCGGCGTGAGACCTACCTGTGTCTAGCTAGTGAAGAGCGGTGACAGTGTCTACA 199
Db	61	GCTAAGGGTCGGCGTGAGACCTACCTGTGTCTAGCTAGTGAAGAGCGGTGACAGTGTCTACA 120
QY	200	TCCTTTTCACTGGACCTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 259
Db	121	TCCTTTTCACTGGACCTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 180
QY	260	TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGTCTACCGGTCACCTGG 319
Db	181	TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGTCTACCGGTCACCTGG 240
QY	320	TTCAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAAAGAACTTTCAAA 379
Db	241	TTCAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAAAGAACTTTCAAA 300
QY	380	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
Db	301	AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
QY	440	GCTGAGCCCGAGGGCTGGCGGCGGTGACCGCGCGGGGTGCAATAGCCATCATGACC 499
Db	361	GCTGAGCCCGAGGGCTGGCGGCGGTGACCGCGCGGGGTGCAATAGCCATCATGACC 420
QY	500	TTCAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAAAGAACTTTCAAA 559
Db	421	TTCAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAAAGAACTTTCAAA 480
QY	560	GCTGGGAAGGGCTGCATGAAATTCAGTTCCAGACAGCTTCGACTTTGGGACTTTGA 619
Db	481	GCTGGGAAGGGCTGCATGAAATTCAGTTCCAGACAGCTTCGACTTTGGGACTTTGA 540
QY	620	TTGCCCTCTATGAGGTTGATGACTTACGAGACGCAATTTTCGTACTTTGGGACTTTG 675

```
|||||
Db 541 TTGCCCCCTGATGAGTTGATGACTTACGAGCGCATTTTCGTACTTTGGGACTTTG 596
|||||

RESULT 8
AF529816
LOCUS Homo sapiens clone Ramos 2 AID (AID) mRNA linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 2 AID (AID) mRNA, partial cds.
ACCESSION AF529816
VERSION AF529816.1 GI:22297219
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Martin.A. and Scharff,M.D.
JOURNAL Somatic hypermutation of the AID transgene in B cells and non-B
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
JOURNAL Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 2"
1..>596
/gene="AID"
1..>596
/gene="AID"
1..>596
/feature="integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95403.1"
/db_xref="GI:22297220"
/translation="MDSLNNRRKFLYQFKNVKAKGRRETYLCYVVKRRDSATSFSL
DGLYLRNKGCHVELLFLRYISDWLDPGRYRVTFWTSRSPCYDCARHVAFLRGNP
NLSLRIFARLYFCEDRKAEPGLRHLRHAGVQIAIMTFKDYFCWNTFVENHERTEFK
AWEGHLHNSVRLSRQLRILLPLYEVDLDRDARTLGL"
BASE COUNT 127 a 165 c 155 g 149 t
ORIGIN
Query Match 21.1%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 6.5e-107;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCCTGG 60
QY 140 GCTAAGGGTCGGCGTGAGACCTACCTGTCTAGTAGTGAAGAGGGGTGACAGTGTCTACA 199
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTCTAGTAGTGAAGAGGGGTGACAGTGTCTACA 120
QY 200 TCTTTTCTACTGGACTTTCGTTATCTTCGCAATAAGAACGGGTGCCACGTGGAATTCGTC 259
Db 121 TCTTTTCTACTGGACTTTCGTTATCTTCGCAATAAGAACGGGTGCCACGTGGAATTCGTC 180
QY 260 TTCTCCCGCTACATCTCGGACCTGGGACCTAGACCTGGCGCGTGTACCGCGTCACTCGG 319
Db 181 TTCTCCCGCTACATCTCGGACCTGGGACCTAGACCTGGCGCGTGTACCGCGTCACTCGG 240
QY 320 TTCACTCTCTGGAGCCCTCTGCTACGACTGTGCCGACATGTGGCCGACTTCTTTCGAGGG 379
Db 241 TTCACTCTCCGGAGCCCTCTGCTACGACTGTGCCGACATGTGGCCGACTTCTTTCGAGGG 300
QY 380 AACCCCAACCTCAGTGTAGGATCTTTCACCGCGCGCTCTACTCTGTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTGTAGGATCTTTCACCGCGCGCTCTACTCTGTGTGAGGACCGCAAG 360
|||||
Db 440 GCTGAGCCCGAGGGCTGCGGGCTGCACCGCGCGGGTGCAGAAATAGCCATCATGACC 499
|||||
Db 361 GCTGAGCCCGAGGGCTGCGGGCTGCACCGCGCGGGTGCAGAAATAGCCATCATGACC 420
|||||
QY 500 TTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAAGACTTTCAAA 559
|||||
Db 421 TTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAAGACTTTCAAA 480
|||||
QY 560 GCTGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 619
|||||
Db 481 GCTGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
|||||
QY 620 TTGCCCCCTGATGAGTTGATGACTTACGAGCGCATTTTCGTACTTTGGGACTTTG 675
|||||
Db 541 TTGCCCCCTGATGAGTTGATGACTTACGAGCGCATTTTCGTACTTTGGGACTTTG 596
|||||

RESULT 9
AF529819
LOCUS Homo sapiens clone Ramos 5 AID (AID) mRNA linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 5 AID (AID) mRNA, partial cds.
ACCESSION AF529819
VERSION AF529819.1 GI:22297225
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE Martin.A. and Scharff,M.D.
JOURNAL Somatic hypermutation of the AID transgene in B cells and non-B
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin.A. and Scharff,M.D.
TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
JOURNAL Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 5"
1..>596
/gene="AID"
1..>596
/gene="AID"
1..>596
/feature="integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95406.1"
/db_xref="GI:22297226"
/translation="MDSLNNRRKFLYQFKNVKAKGRRETYLCYVVKRRDSATSFSL
DGLYLRNKGCHVELLFLRYISDWLDPGRYRVTFWTSRSPCYDCARHVAFLRGNP
NLSLRIFARLYFCEDRKAEPGLRHLRHAGVQIAIMTFKDYFCWNTFVENHERTEFK
AWEGHLHNSVRLSRQLRILLPLYEVDLDRDARTLGL"
BASE COUNT 128 a 164 c 154 g 150 t
ORIGIN
Query Match 21.1%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 6.5e-107;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCCTGG 139
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGTCGCCTGG 60
QY 140 GCTAAGGGTCGGCGTGAGACCTACCTGTCTAGTAGTGAAGAGGGGTGACAGTGTCTACA 199
Db 61 GCTAAGGGTCGGCGTGAGACCTACCTGTCTAGTAGTGAAGAGGGGTGACAGTGTCTACA 120
QY 200 TCTTTTCTACTGGACTTTCGTTATCTTCGCAATAAGAACGGGTGCCACGTGGAATTCGTC 259
|||||
```



```
Db 121 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 180
Qy 260 TTCCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGGTCACTCGG 319
Db 181 TTCCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGGTCACTCGG 240
Qy 320 TTCACCTTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGGGAGG 379
Db 241 TTCACCTTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGGGAGG 300
Qy 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Qy 440 GCTGAGCCGAGGGCTGGCGGCTGTCACCGCGCGCTGTCACCAATAGAACCGCTGCCACGTGGAATTGCTC 259
Db 361 GCTGAGCCGAGGGCTGGCGGCTGTCACCGCGCGCTGTCACCAATAGAACCGCTGCCACGTGGAATTGCTC 180
Qy 500 TTCACCTTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGGGAGG 319
Db 421 TTCACCTTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGGGAGG 379
Qy 560 GCTGAGCCGAGGGCTGGCGGCTGTCACCGCGCGCTGTCACCAATAGAACCGCTGCCACGTGGAATTGCTC 259
Db 481 GCTGAGCCGAGGGCTGGCGGCTGTCACCGCGCGCTGTCACCAATAGAACCGCTGCCACGTGGAATTGCTC 180
Qy 620 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTTCTGACTTTGGGACTTTG 675
Db 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTTCTGACTTTGGGACTTTG 596

RESULT 10
AF529820
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 596)
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
Location/Qualifiers
1. .596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 6"
1. .>596
/gene="AID"
1. .>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95407.1"
/db_xref="GI:22297228"
/translation="NDSLLMRKFLYQFKNVRWAKGRRETYLCYVKKRDSATSFSL
DFGLRNKNGCHELLFLRYISDWLDPGRYRVWFTSWSPCYDCARHADVFLRGNP
NLSLRTFLRYFCVDRKAEPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHRTFK
AWEGHLHNSRQLRLLPLLYEVDLDRAFRLGL"
126 a 164 c 155 g 151 t

BASE COUNT
ORIGIN
gene
1. .>596

Query Match 21.1%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 6.5e-107;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 80 ATGACACCCCTCTTGATGTAACCGGAGGAAGTTTCTTTACCAATTTCAAAATTTGCCGCTGG 139
Db 1 ATGACACCCCTCTTGATGTAACCGGAGGAAGTTTCTTTACCAATTTCAAAATTTGCCGCTGG 60
Qy 140 GCTAAGGCTCGCGCTGAGACCTACCTGCTGCTACGTAGTGAAGAGCGCTGACAGTGTCTACA 199
Db 61 GCTAAGGCTCGCGCTGAGACCTACCTGCTGCTACGTAGTGAAGAGCGCTGACAGTGTCTACA 120
Qy 200 TCCCTTTTCACTGGAGCTTTGGTTATCTTCGCAATAAGAACCGCTGCCACGTGGAATTGCTC 259
Db 121 TCCCTTTTCACTGGAGCTTTGGTTATCTTCGCAATAAGAACCGCTGCCACGTGGAATTGCTC 180
Qy 260 TTCCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGGTCACTCGG 319
Db 181 TTCCTCGGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGGTCACTCGG 240
Qy 320 TTCACCTTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGGGAGG 379
Db 241 TTCACCTTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGGGAGG 300
Qy 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 439
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Qy 440 GCTGAGCCGAGGGCTGGCGGCTGTCACCGCGCGCTGTCACCAATAGAACCGCTGCCACGTGGAATTGCTC 259
Db 361 GCTGAGCCGAGGGCTGGCGGCTGTCACCGCGCGCTGTCACCAATAGAACCGCTGCCACGTGGAATTGCTC 180
Qy 500 TTCACCTTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGGGAGG 319
Db 421 TTCACCTTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGGGAGG 379
Qy 560 GCTGAGCCGAGGGCTGGCGGCTGTCACCGCGCGCTGTCACCAATAGAACCGCTGCCACGTGGAATTGCTC 259
Db 481 GCTGAGCCGAGGGCTGGCGGCTGTCACCGCGCGCTGTCACCAATAGAACCGCTGCCACGTGGAATTGCTC 180
Qy 620 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTTCTGACTTTGGGACTTTG 675
Db 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTTCTGACTTTGGGACTTTG 596

RESULT 11
AF529821
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 596)
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
Location/Qualifiers
1. .596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 7"
1. .>596
/gene="AID"
1. .>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95407.1"
/db_xref="GI:22297228"
/translation="NDSLLMRKFLYQFKNVRWAKGRRETYLCYVKKRDSATSFSL
DFGLRNKNGCHELLFLRYISDWLDPGRYRVWFTSWSPCYDCARHADVFLRGNP
NLSLRTFLRYFCVDRKAEPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHRTFK
AWEGHLHNSRQLRLLPLLYEVDLDRAFRLGL"
126 a 164 c 155 g 151 t

BASE COUNT
ORIGIN
gene
1. .>596

AF529821
Homo sapiens clone Ramos 7 AID (AID) mRNA, partial cds.
AF529821.1 GI:22297229
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 596)
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
Location/Qualifiers
1. .596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 7"
1. .>596
/gene="AID"
1. .>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95407.1"
/db_xref="GI:22297228"
/translation="NDSLLMRKFLYQFKNVRWAKGRRETYLCYVKKRDSATSFSL
DFGLRNKNGCHELLFLRYISDWLDPGRYRVWFTSWSPCYDCARHADVFLRGNP
NLSLRTFLRYFCVDRKAEPEGLRLHRAGVQIAIMTFKDYFCWNTFVENHRTFK
AWEGHLHNSRQLRLLPLLYEVDLDRAFRLGL"
126 a 164 c 155 g 151 t

BASE COUNT
ORIGIN
gene
1. .>596
```



CDS	/gene="AID" 1. .>596 /gene="AID" /note="Integrated into Burkitt's lymphoma cell line Ramos" /codon_start=1 /product="AID" /protein_id="AA095408.1" /db_xref="GI:22297230" /translation="MDSLMMRRKFLYQFNVRWAKGRRETYLCYVVKRRDRATSFSL DFGYLRNKGCHVELLFYISDMDPDGRCYRVITWFTSPCYDCARHVADEFGRNP NLSLRIETARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYFCWNTFEVENHERITPK ANEGLHENSVRSLRQLRLLPLLYEVDLDRDAFTLGL" BASE COUNT 127 a 165 c 155 g 149 t ORIGIN		/gene="AID" 1. .>596 /gene="AID" /note="Integrated into Burkitt's lymphoma cell line Ramos" /db_xref="taxon:9606" /clone="Ramos 8" 1. .>596 /gene="AID" /gene="AID" /note="Integrated into Burkitt's lymphoma cell line Ramos" /codon_start=1 /product="AID" /protein_id="GI:22297232" /db_xref="GI:22297232" /translation="MDSLMMRRKFLYQFNVRWAKGRRETYLCYVVKRRDRATSFSL DFGYLRNKGCHVELLFYISDMDPDGRCYRVITWFTSPCYDCARHVADEFGRNP NLSLRIETARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYFCWNTFEVENHERITPK ANEGLHENSVRSLRQLRLLPLLYEVDLDRDAFTLGL" BASE COUNT 126 a 165 c 155 g 150 t ORIGIN		Query Match 21.1%; Score 594.4; DB 9; Length 596; Best Local Similarity 99.8%; Pred. No. 6.5e-107; Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;		Query Match 21.1%; Score 594.4; DB 9; Length 596; Best Local Similarity 99.8%; Pred. No. 6.5e-107; Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
	Qy	80	ATGCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGGCTGG	139	Qy	80	ATGCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGGCTGG	139
	Db	1	ATGCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGGCTGG	60	Db	1	ATGCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGGCTGG	60
	Qy	140	GCTAAGGGTCGGCGTGACACCTACCTGCTACGTAGTGAAGAGCGGTGACAGTGCCTACA	199	Qy	140	GCTAAGGGTCGGCGTGACACCTACCTGCTACGTAGTGAAGAGCGGTGACAGTGCCTACA	199
	Db	61	GCTAAGGGTCGGCGTGACACCTACCTGCTACGTAGTGAAGAGCGGTGACAGTGCCTACA	120	Db	61	GCTAAGGGTCGGCGTGACACCTACCTGCTACGTAGTGAAGAGCGGTGACAGTGCCTACA	120
	Qy	200	TCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGAACGGTGCACGTGGAATTGCTC	259	Qy	200	TCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGAACGGTGCACGTGGAATTGCTC	259
	Db	121	TCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGAACGGTGCACGTGGAATTGCTC	180	Db	121	TCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGAACGGTGCACGTGGAATTGCTC	180
	Qy	260	TTCCCTCCGCTACATCTCGGACTGGACCTAGACCTGCGCCGCTGCTACCGGTCCACCTGG	319	Qy	260	TTCCCTCCGCTACATCTCGGACTGGACCTAGACCTGCGCCGCTGCTACCGGTCCACCTGG	319
	Db	181	TTCCCTCCGCTACATCTCGGACTGGACCTAGACCTGCGCCGCTGCTACCGGTCCACCTGG	240	Db	181	TTCCCTCCGCTACATCTCGGACTGGACCTAGACCTGCGCCGCTGCTACCGGTCCACCTGG	240
	Qy	320	TTCCACCTCTCGAGCGCTGCTACGACTGTGCGCCGACATGTGGCGGCTTTCTCGAGGG	379	Qy	320	TTCCACCTCTCGAGCGCTGCTACGACTGTGCGCCGACATGTGGCGGCTTTCTCGAGGG	379
	Db	241	TTCCACCTCTCGAGCGCTGCTACGACTGTGCGCCGACATGTGGCGGCTTTCTCGAGGG	300	Db	241	TTCCACCTCTCGAGCGCTGCTACGACTGTGCGCCGACATGTGGCGGCTTTCTCGAGGG	300
	Qy	380	AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGAGACCGCAAG	439	Qy	380	AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGAGACCGCAAG	439
	Db	301	AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGAGACCGCAAG	360	Db	301	AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGAGACCGCAAG	360
	Qy	440	GCTGAGCGCGAGGGCTGCGCGGCTGCGCCGCTGCGCCGCTGCGCCGCTGCGCCGCTGCG	499	Qy	440	GCTGAGCGCGAGGGCTGCGCGGCTGCGCCGCTGCGCCGCTGCGCCGCTGCGCCGCTGCG	499
	Db	361	GCTGAGCGCGAGGGCTGCGCGGCTGCGCCGCTGCGCCGCTGCGCCGCTGCGCCGCTGCG	420	Db	361	GCTGAGCGCGAGGGCTGCGCGGCTGCGCCGCTGCGCCGCTGCGCCGCTGCGCCGCTGCG	420
	Qy	500	TTCAAGAGATTTATTTTACTGCTGGAATATCTTTGTAGAAAACCATGAAAGAACTTTTCAA	559	Qy	500	TTCAAGAGATTTATTTTACTGCTGGAATATCTTTGTAGAAAACCATGAAAGAACTTTTCAA	559
	Db	421	TTCAAGAGATTTATTTTACTGCTGGAATATCTTTGTAGAAAACCATGAAAGAACTTTTCAA	480	Db	421	TTCAAGAGATTTATTTTACTGCTGGAATATCTTTGTAGAAAACCATGAAAGAACTTTTCAA	480
	Qy	560	GCTGGGAAGGGCTGCATGAAATTCAGTTTCCTCCAGACAGCTTCGCGCGCATCCTT	619	Qy	560	GCTGGGAAGGGCTGCATGAAATTCAGTTTCCTCCAGACAGCTTCGCGCGCATCCTT	619
	Db	481	GCTGGGAAGGGCTGCATGAAATTCAGTTTCCTCCAGACAGCTTCGCGCGCATCCTT	540	Db	481	GCTGGGAAGGGCTGCATGAAATTCAGTTTCCTCCAGACAGCTTCGCGCGCATCCTT	540
	Qy	620	TTGCCCTGTATGAGGTTGATGACATTACGAGACGCAATTCGTAATTTGGGACTTTG	675	Qy	620	TTGCCCTGTATGAGGTTGATGACATTACGAGACGCAATTCGTAATTTGGGACTTTG	675
	Db	541	TTGCCCTGTATGAGGTTGATGACATTACGAGACGCAATTCGTAATTTGGGACTTTG	596	Db	541	TTGCCCTGTATGAGGTTGATGACATTACGAGACGCAATTCGTAATTTGGGACTTTG	596
RESULT 12								
AF529822	AF529822.1 GI:22297231							
LOCUS	Homo sapiens clone Ramos 8 AID (AID) mRNA, partial cds.							
DEFINITION	Homo sapiens clone Ramos 8 AID (AID) mRNA, partial cds.							
ACCESSION	AF529822							
VERSION	AF529822.1							
KEYWORDS	human.							
SOURCE	Homo sapiens							
ORGANISM	Homo sapiens							
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.							
AUTHORS	Martin, A. and Scharff, M.D.							

```
RESULT 13
AF529823
LOCUS Homo sapiens clone Ramos 9 AID (AID) mRNA linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 9 AID (AID) mRNA, partial cds.
ACCESSION AF529823
VERSION AF529823.1 GI:22297233
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 9"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/db_xref="GI:22297234"
/translation="MDSLNNRRKFLYQFKNVWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGHVELFLRYISDWLDPGRCYRVWTFWSWSPCYDCARHVDPLRGNP
NLSLRIFTARLYECEDRKAEPGLRLHRAGVQIAIMTFKDYFCWNTFVENHRTFK
AWEGHNSVRLSRQLRILLPLYEVDLDRDAFRTGL"
BASE COUNT 128 a 163 c 155 g 150 t
ORIGIN
Query Match 21.1%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 6.5e-107;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCTGG 139
DB 1 ATGCACAGCCTATTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCTGG 60
QY 140 GCTAAGGGTGGCGGTGAGACCTACCTGCTAGTAGTAGTGAAGAGGCGTGACAGTGCTACA 199
DB 61 GCTAAGGGTGGCGGTGAGACCTACCTGCTAGTAGTAGTGAAGAGGCGTGACAGTGCTACA 120
QY 200 TCCTTTTCACTGGACCTTGGTATCTTCGCAATAAGAACGGCTGCCACGTGGAATGGCTC 259
DB 121 TCCTTTTCACTGGACCTTGGTATCTTCGCAATAAGAACGGCTGCCACGTGGAATGGCTC 180
QY 260 TTCTCTCGGTACATCTCGACCTGGGACCTAGACCTGGCGCTGCTACCGCGTCACTCTGG 319
DB 181 TTCTCTCGGTACATCTCGACCTGGGACCTAGACCTGGCGCTGCTACCGCGTCACTCTGG 240
QY 320 TTCACTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGGCAGGG 379
DB 241 TTCACTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGGCAGGG 300
QY 380 AACCCCAACTCAGTCTGAGGATCTTACCGCGCGCCTCTACTTCTGTGAGACCGCAAG 439
DB 301 AACCCCAACTCAGTCTGAGGATCTTACCGCGCGCCTCTACTTCTGTGAGACCGCAAG 360
QY 440 GCTGAGCCGAGGGGTGCGGGGCTGACCGCGCGGGGTGCAACATAGCCATCATGACC 499
|||||
```

```
Db 361 GCTGAGCCCGAGGGCTGCGGGCGTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
QY 500 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCATCAAGAACTTTTCAAA 559
|||||
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAACCATCAAGAACTTTTCAAA 480
|||||
QY 560 GCCTGGGAAGGGCTGCATGAAAATTCAGTTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 619
|||||
Db 481 GCCTGGGAAGGGCTGCATGAAAATTCAGTTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 540
|||||
QY 620 TTGCCCTCTGATGAGGTGATGACTTACGAGAGCGCATTTTCGTACTTTGGGACTTTG 675
|||||
Db 541 TTGCCCTCTGATGAGGTGATGACTTACGAGAGCGCATTTTCGTACTTTGGGACTTTG 596
|||||

RESULT 14
AF529824
LOCUS Homo sapiens clone Ramos 10 AID (AID) mRNA linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 10 AID (AID) mRNA, partial cds.
ACCESSION AF529824
VERSION AF529824.1 GI:22297235
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 10"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95411.1"
/db_xref="GI:22297236"
/translation="MDSLNNRRKFLYQFKNVWAKGRRETYLCYVVKRRDSATSFSL
DFGYLRNKNHVELFLRYISDWLDPGRCYRVWTFWSWSPCYDCARHVDPLRGNP
NLSLRIFTARLYECEDRKAEPGLRLHRAGVQIAIMTFKDYFCWNTFVENHRTFK
AWEGHNSVRLSRQLRILLPLYEVDLDRDAFRTGL"
BASE COUNT 127 a 163 c 156 g 150 t
ORIGIN
Query Match 21.1%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 6.5e-107;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCTGG 139
|||||
Db 1 ATGCACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAATGTCGCTGG 60
QY 140 GCTAAGGGTGGCGGTGAGACCTACCTGCTAGTAGTAGTGAAGAGGCGTGACAGTGCTACA 199
|||||
Db 61 GCTAAGGGTGGCGGTGAGACCTACCTGCTAGTAGTAGTGAAGAGGCGTGACAGTGCTACA 120
QY 200 TCCTTTTCACTGGACCTTGGTATCTTCGCAATAAGAACGGCTGCCACGTGGAATGGCTC 259
|||||
Db 121 TCCTTTTCACTGGACCTTGGTATCTTCGCAATAAGAACGGCTGCCACGTGGAATGGCTC 180
|||||
```

```
QY 260 TTCCTCCGCTACATCGGACTGCGACCTAGACCCTGGCCGCTGCTACCGCTCACCTGG 319
    |||||
Db 181 TTCCTCCGCTACATCGGACTGCGACCTAGACCCTGGCCGCTGCTACCGCTCACCTGG 240
QY 320 TTCACCTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTCTGGAGGG 379
    |||||
Db 241 TTCACCTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTCTGGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCCTCTACTTCTGTGAGGACCGCAAG 439
    |||||
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCGAGGGGCTGCGCGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 499
    |||||
Db 361 GCTGAGCCGAGGGGCTGCGCGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
QY 500 TTCAAAGATATTTTACTGCTGGAATTCAGTTCGCTCTCTCCAGACAGCTTTCGCGCATCCTT 559
    |||||
Db 421 TTCAAAGATATTTTACTGCTGGAATTCAGTTCGCTCTCTCCAGACAGCTTTCGCGCATCCTT 480
QY 560 GCCTGGGAGGGCTGCATGAAATTCAGTTCGCTCTCTCCAGACAGCTTTCGCGCATCCTT 619
    |||||
Db 481 GCCTGGGAGGGCTGCATGAAATTCAGTTCGCTCTCTCCAGACAGCTTTCGCGCATCCTT 540
QY 620 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGCTACTTTGGGACTTTG 675
    |||||
Db 541 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGCTACTTTGGGACTTTG 596
```

```
RESULT 15
AF529825      596 bp      mRNA      linear      PRI 19-AUG-2002
LOCUS      Homo sapiens clone Ramos 11 AID (AID) mRNA, partial cds.
DEFINITION      AF529825
ACCESSION      AF529825
VERSION      AF529825.1 GI:22297237
KEYWORDS
SOURCE      human.
ORGANISM      Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 596)
AUTHORS      Martin,A. and Scharff,M.D.
TITLE      Somatic hypermutation of the AID transgene in B cells and non-B
               cells
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 596)
AUTHORS      Martin,A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL      Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
               Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
               Location/Qualifiers
               1..596
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="Ramos 11"
               1..>596
               /gene="AID"
               1..>596
               /gene="AID"
               /note="Integrated into Burkitt's lymphoma cell line Ramos"
               /codon_start=1
               /product="AID"
               /protein_id="AA095412.1"
               /db_xref="GI:22297238"
               /translation="MDSLMMNRKFLYQKNVRWAKRREAYLCVYVKKRDSATSFSL
               DFGYLRNKGHVELLFYRISDWDLDFRCYRVTFWTSWSPCYDCARHVAFLRGNP
               NLSRIFARLYFCEDRKAEPGLRHLRAGVOIAIMTFKDYFYCWNTFVENHETFK
               AWEGLHENSRLRLLPLLYEVDLDRARFTLGL"
BASE COUNT      126 a 164 c 156 g 150 t
ORIGIN
```

Query Match 21.1%; Score 594.4; DB 9; Length 596;  
Best Local Similarity 99.8%; Pred. No. 6.5e-107;

```
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 80 ATGGACAGCCTCTTATGAACCGGAGGAGTTTCTTTACCAATTTCAAAATCTCGCGTGG 139
    |||||
Db 1 ATGGACAGCCTCTTATGAACCGGAGGAGTTTCTTTACCAATTTCAAAATCTCGCGTGG 60
QY 140 GCTAAGGCTGGCGGTGAGACCTTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTACA 199
    |||||
Db 61 GCTAAGGCTGGCGGTGAGACCTTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTACA 120
QY 200 TCCTTTTTCACGTGAGACTTTGGTTATCTTCGCAATAAGACGGGTGCACGTGGAATGTCTC 259
    |||||
Db 121 TCCTTTTTCACGTGAGACTTTGGTTATCTTCGCAATAAGACGGGTGCACGTGGAATGTCTC 180
QY 260 TTCCCTCCGCTACATCTCGGACTTGGGACCTAGACCTTGGCCGCTGTACCGCGTCACTCGG 319
    |||||
Db 181 TTCCCTCCGCTACATCTCGGACTTGGGACCTAGACCTTGGCCGCTGTACCGCGTCACTCGG 240
QY 320 TTCACCTCCTGGAGCCCTGCTACGACTGTGCTCCGACATGTGGCCGACTTCTTGGGAGGG 379
    |||||
Db 241 TTCACCTCCTGGAGCCCTGCTACGACTGTGCTCCGACATGTGGCCGACTTCTTGGGAGGG 300
QY 380 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCCTCTACTTCTGTGAGGACCGCAAG 439
    |||||
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCCTCTACTTCTGTGAGGACCGCAAG 360
QY 440 GCTGAGCCGAGGGGCTGCGCGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 499
    |||||
Db 361 GCTGAGCCGAGGGGCTGCGCGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
QY 500 TTCAAAGATATTTTACTGCTGGAATTCAGTTCGCTCTCTCCAGACAGCTTTCGCGCATCCTT 559
    |||||
Db 421 TTCAAAGATATTTTACTGCTGGAATTCAGTTCGCTCTCTCCAGACAGCTTTCGCGCATCCTT 480
QY 560 GCCTGGGAGGGCTGCATGAAATTCAGTTCGCTCTCTCCAGACAGCTTTCGCGCATCCTT 619
    |||||
Db 481 GCCTGGGAGGGCTGCATGAAATTCAGTTCGCTCTCTCCAGACAGCTTTCGCGCATCCTT 540
QY 620 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGCTACTTTGGGACTTTG 675
    |||||
Db 541 TTGCCCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGCTACTTTGGGACTTTG 596
```

Search completed: June 19, 2003, 07:46:30  
Job time : 7197.7 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 23:33:05 ; Search time 1524.43 Seconds  
(without alignments)  
11397.322 Million cell updates/sec

Title: US-09-966-880A-7\_COPY\_80\_676  
Perfect score: 597  
Sequence: 1 atggacagcctcttgatgaa.....ttcgactttgggactttga 597

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*

Result No.	Score	Query Match	Length	DB ID	Description
1	597	100.0	1837	9	BC006296 Homo sapi
2	597	100.0	2791	9	AB040431 Homo sapi
3	595.4	99.7	597	12	AF529828 Mus muscu
4	594.4	99.6	596	9	AF529815 Homo sapi
5	594.4	99.6	596	9	AF529816 Homo sapi
6	594.4	99.6	596	9	AF529819 Homo sapi
7	594.4	99.6	596	9	AF529820 Homo sapi
8	594.4	99.6	596	9	AF529821 Homo sapi
9	594.4	99.6	596	9	AF529822 Homo sapi
10	594.4	99.6	596	9	AF529823 Homo sapi
11	594.4	99.6	596	9	AF529824 Homo sapi
12	594.4	99.6	596	9	AF529825 Homo sapi
13	594.4	99.6	596	9	AF529826 Homo sapi
14	594.4	99.6	596	9	AF529827 Homo sapi
15	594.4	99.6	596	12	AF529829 Mus muscu
16	594.4	99.6	596	12	AF529830 Mus muscu
17	594.4	99.6	596	12	AF529831 Mus muscu
18	594.4	99.6	596	12	AF529833 Mus muscu
19	594.4	99.6	596	12	AF529835 Mus muscu
20	594.4	99.6	596	12	AF529837 Mus muscu
21	594.4	99.6	596	12	AF529839 Mus muscu
22	593.4	99.4	595	12	AF529834 Mus muscu
23	592.8	99.3	596	9	AF529817 Homo sapi
24	592.8	99.3	596	9	AF529818 Homo sapi
25	592.8	99.3	596	12	AF529832 Mus muscu
26	592.8	99.3	596	12	AF529836 Mus muscu
27	592.8	99.3	596	12	AF529838 Mus muscu
28	592.2	99.2	597	12	AF529841 Cricetulu
29	592.2	99.2	597	12	AF529842 Cricetulu
30	592.2	99.2	597	12	AF529843 Cricetulu
31	592.2	99.2	597	12	AF529844 Cricetulu
32	592.2	99.2	597	12	AF529845 Cricetulu
33	592.2	99.2	597	12	AF529846 Cricetulu
34	592.2	99.2	597	12	AF529847 Cricetulu
35	592.2	99.2	597	12	AF529848 Cricetulu
36	592.2	99.2	597	12	AF529849 Cricetulu
37	592.2	99.2	597	12	AF529850 Cricetulu
38	592.2	99.2	597	12	AF529851 Cricetulu
39	592.2	99.2	597	12	AF529852 Cricetulu
40	592.2	99.2	597	12	AF529853 Cricetulu
41	592.2	99.2	597	12	AF529854 Cricetulu
42	586.2	98.2	591	12	AF529855 Cricetulu
43	575.4	96.4	577	12	AF529840 Mus muscu
44	542.2	90.8	547	12	AF529856 Cricetulu
45	475.4	79.6	2440	10	AF132979 Mus muscu

ALIGNMENTS

RESULT 1  
BC006296  
LOCUS BC006296  
DEFINITION Homo sapiens, activation-induced cytidine deaminase, clone  
MGC:12911 IMAGE:4054915, mRNA, complete cds.  
ACCESSION BC006296  
VERSION MGC.  
KEYWORDS MGC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1837)  
AUTHORS Strausberg,R.  
TITLE Direct Submission

Pred. No. is the number of results predicted by chance to have a



```
BASE COUNT      842 a   548 c   625 g   776 t

Query Match      100.0%; Score 597; DB 9; Length 2791;
Best Local Similarity 100.0%; Pred. No. 3.3e-150;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGACAGCCTCTTGATGAACCGGAGAGAGTTTCTTTACCAATTCAAAAATGCCGCTGG 60
   |||||||
Db 77 ATGGACAGCCTCTTGATGAACCGGAGAGAGTTTCTTTACCAATTCAAAAATGCCGCTGG 136

QY 61 GCTAAGGTCGGCGTGAGACCTACCTGTGCTAGTAGTGAAGAGCGGTGACAGTGCTACA 120
   |||||||
Db 137 GCTAAGGTCGGCGTGAGACCTACCTGTGCTAGTAGTGAAGAGCGGTGACAGTGCTACA 196

QY 121 TCCCTTTTCACTGACCTTTGGTTATCTTCGCAATAGAAGCGGTGCCACGTGGAATGGCTC 180
   |||||||
Db 197 TCCCTTTTCACTGACCTTTGGTTATCTTCGCAATAGAAGCGGTGCCACGTGGAATGGCTC 256

QY 181 TTCCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGTCTACCGCGCTACCTGG 240
   |||||||
Db 257 TTCCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGTCTACCGCGCTACCTGG 316

QY 241 TTCACCTCTCGAGCCCTGTGCTAGCAGCTGTGCCCGACATGTGCCCGACCTTTCTGCGAGGG 300
   |||||||
Db 317 TTCACCTCTCGAGCCCTGTGCTAGCAGCTGTGCCCGACATGTGCCCGACCTTTCTGCGAGGG 376

QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
   |||||||
Db 377 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 436

QY 361 GCTGAGCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 420
   |||||||
Db 437 GCTGAGCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 496

QY 421 TTCAAAGATATTTTACTGCTGGAATACATTTTGTAGAAAACATGAAGAACTTTTCAA 480
   |||||||
Db 497 TTCAAAGATATTTTACTGCTGGAATACATTTTGTAGAAAACATGAAGAACTTTTCAA 556

QY 481 GCTGGGAAGGCTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGCATCTT 540
   |||||||
Db 557 GCTGGGAAGGCTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGCATCTT 616

QY 541 TTGCCCTCTATGAGTTTGATGACTTACGAGACGATTTTCGATTTGGGACTTTGA 597
   |||||||
Db 617 TTGCCCTCTATGAGTTTGATGACTTACGAGACGATTTTCGATTTGGGACTTTGA 673

RESULT 3
AF529828      597 bp      mRNA      linear      SYN 19-AUG-2002
LOCUS      Mus musculus clone 1 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION      complete cds.
ACCESSION      AF529828
VERSION      AF529828.1 GI:22297243
KEYWORDS
SOURCE      house mouse.
ORGANISM      Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE      1 (bases 1 to 597)
AUTHORS      Martin.A. and Scharff,M.D.
TITLE      Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL      Unpublished
REFERENCE      2 (bases 1 to 597)
AUTHORS      Martin.A. and Scharff,M.D.
TITLE      Direct Submission
JOURNAL      Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
               Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..597
/organism="Mus musculus"
/db_xref="taxon:10090"

/clone="1"
/cell_type="hybridoma Pl-5"
/transgenic
1..597
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..597
/gene="AID"
1..597
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95415.1"
/db_xref="GI:2229724"
/translation="MDSLNNRRKFLYOFKNRWAKGRRETYLCYVVKRRDSATPSL
DRGYLRNKNDCHVELLFLEYISDWLDPGRCYRVTFWTSFSCYDCARHVADEFLRNP
NLSLRFTFARLYFCEDRKAEPGLRHRAGVOIAIMTFRDIYFCWNTFVENHRTFFK
AWEGJHNSVRLSRQLRLLPLLYEVDLDRDAFRTGL"
BASE COUNT      129 a   164 c   154 g   150 t
ORIGIN

Query Match      99.7%; Score 595.4; DB 12; Length 597;
Best Local Similarity 99.8%; Pred. No. 7e-150;
Matches 596; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGACAGCCTCTTGATGAACCGGAGAGAGTTTCTTTACCAATTCAAAAATGCCGCTGG 60
   |||||||
Db 1 ATGGACAGCCTCTTGATGAACCGGAGAGAGTTTCTTTACCAATTCAAAAATGCCGCTGG 60

QY 61 GCTAAGGTCGGCGTGAGACCTACCTGTGCTAGTAGTGAAGAGCGGTGACAGTGCTACA 120
   |||||||
Db 61 GCTAAGGTCGGCGTGAGACCTACCTGTGCTAGTAGTGAAGAGCGGTGACAGTGCTACA 120

QY 121 TCCCTTTTCACTGACCTTTGGTTATCTTCGCAATAGAAGCGGTGCCACGTGGAATGGCTC 180
   |||||||
Db 121 TCCCTTTTCACTGACCTTTGGTTATCTTCGCAATAGAAGCGGTGCCACGTGGAATGGCTC 180

QY 181 TTCCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGTCTACCGCGCTACCTGG 240
   |||||||
Db 181 TTCCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGTCTACCGCGCTACCTGG 240

QY 241 TTCACCTCTCGAGCCCTGTGCTAGCAGCTGTGCCCGACATGTGCCCGACCTTTCTGCGAGGG 300
   |||||||
Db 241 TTCACCTCTCGAGCCCTGTGCTAGCAGCTGTGCCCGACATGTGCCCGACCTTTCTGCGAGGG 300

QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
   |||||||
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360

QY 361 GCTGAGCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 420
   |||||||
Db 361 GCTGAGCCGAGGGGTGCGGGGCTGCACCGCGCGGGTGCAAATAGCCATCATGACC 420

QY 421 TTCAAAGATATTTTACTGCTGGAATACATTTTGTAGAAAACATGAAGAACTTTTCAA 480
   |||||||
Db 421 TTCAAAGATATTTTACTGCTGGAATACATTTTGTAGAAAACATGAAGAACTTTTCAA 480

QY 481 GCTGGGAAGGCTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGCATCTT 540
   |||||||
Db 481 GCTGGGAAGGCTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGCATCTT 540

QY 541 TTGCCCTCTATGAGTTTGATGACTTACGAGACGATTTTCGATTTGGGACTTTGA 597
   |||||||
Db 541 TTGCCCTCTATGAGTTTGATGACTTACGAGACGATTTTCGATTTGGGACTTTGA 597

RESULT 4
AF529815
LOCUS      Homo sapiens clone Ramos 1 AID (AID) mRNA, partial cds.
DEFINITION      AF529815
ACCESSION      AF529815
VERSION      AF529815.1 GI:22297217
KEYWORDS
```

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE 1 (bases 1 to 596)  
JOURNAL Martin, A. and Scharff, M.D.  
FEATURES Somatic hypermutation of the AID transgene in B cells and non-B  
cells  
Unpublished  
2 (bases 1 to 596)  
REFERENCE Martin, A. and Scharff, M.D.  
AUTHORS Direct Submission  
TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
JOURNAL Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES Location/Qualifiers  
source  
1..596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Ramos 1"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA095402.1"  
/db\_xref="GI:22297218"  
/translation="MDSLMMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL  
DFGYLRNKGCHVELLFLRYISDMDDPGRCYRVTFWTSWSPCYDCAHVAADFURGNP  
NLSLRIFARLYFCEDRAEPEGLRLHRAQVQIAIMTFKDYFCWNTFVENHETERK  
AWEGLEHNSVRLSRIRILLPLVEVDLDRDAFRITGL"  
BASE COUNT 128 a 163 c 155 g 150 t  
ORIGIN  
Query Match 99.6%; Score 594.4; DB 9; Length 596;  
Best Local Similarity 99.8%; Pred. No. 1.3e-149;  
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGACACGCTCTTGATGAACCGGAGGAGTTCCTTACCAATTCAAAATGTCGCTGG 60  
DB |  
1 ATGACACGCTCTTGATGAACCGGAGGAGTTCCTTACCAATTCAAAATGTCGCTGG 60  
QY 61 GCTAAGGTCGCGGTGAGACCTACCTGTGTACGTAGAGAGCGGTGACAGTGGCTC 120  
DB |  
61 GCTAAGGTCGCGGTGAGACCTACCTGTGTACGTAGAGAGCGGTGACAGTGGCTC 120  
QY 121 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAACGCGTGCACAGTGGATTGCTC 180  
DB |  
121 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAACGCGTGCACAGTGGATTGCTC 180  
QY 181 TTCCTCCGCTACATCTCGGACTGGACCTAGACCCCTGGCGGTGCTACCGCGTCACTGG 240  
DB |  
181 TTCCTCCGCTACATCTCGGACTGGACCTAGACCCCTGGCGGTGCTACCGCGTCACTGG 240  
QY 241 TTCACCTCTCGGAGCCCTCTGCTACGACTGTGCGCGGCTGCTACCGCGTCACTGG 300  
DB |  
241 TTCACCTCTCGGAGCCCTCTGCTACGACTGTGCGCGGCTGCTACCGCGTCACTGG 300  
QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTCTGTGAGAACCGCAAG 360  
DB |  
301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTCTGTGAGAACCGCAAG 360  
QY 361 GCTGAGCCGAGGGGTGCGGGGTGACACCGCGCGGGGTGCAAAATAGCCATCATGACC 420  
DB |  
361 GCTGAGCCGAGGGGTGCGGGGTGACACCGCGCGGGGTGCAAAATAGCCATCATGACC 420  
QY 421 TTCAAAGATTATTTTACTGCTGGAAATCTTTGTAGAAACCATGAAAGACTTTTCAA 480  
DB |  
421 TTCAAAGATTATTTTACTGCTGGAAATCTTTGTAGAAACCATGAAAGACTTTTCAA 480  
QY 481 GCCTGGAGGCTCGATGAAATTCAGTTGCTCTCTCCAGACAGCTTCGGCGCATCCTT 540  
DB |  
481 GCCTGGAGGCTCGATGAAATTCAGTTGCTCTCTCCAGACAGCTTCGGCGCATCCTT 540

Db 481 GCCTGGAGGCTCGATGAAATTCAGTTGCTCTCTCCAGACAGATTCGGCGCATCCTT 540  
QY 541 TTGCCCCCTATGACGTTGATGACCTAGACAGCAGCATTTGCTACTTTGGGACTTTG 596  
DB |  
541 TTGCCCCCTATGACGTTGATGACCTAGACAGCAGCATTTGCTACTTTGGGACTTTG 596  
RESULT 5  
AF529816 596 bp mRNA linear PRI 19-AUG-2002  
LOCUS Homo sapiens clone Ramos 2 AID (AID) mRNA, partial cds.  
DEFINITION  
ACCESSION AF529816  
VERSION AF529816.1 GI:22297219  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE 1 (bases 1 to 596)  
JOURNAL Martin, A. and Scharff, M.D.  
FEATURES Somatic hypermutation of the AID transgene in B cells and non-B  
cells  
Unpublished  
2 (bases 1 to 596)  
REFERENCE Martin, A. and Scharff, M.D.  
AUTHORS Direct Submission  
TITLE Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
JOURNAL Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES Location/Qualifiers  
source  
1..596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Ramos 2"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AA095403.1"  
/db\_xref="GI:22297220"  
/translation="MDSLMMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSL  
DFGYLRNKGCHVELLFLRYISDMDDPGRCYRVTFWTSWSPCYDCAHVAADFURGNP  
NLSLRIFARLYFCEDRAEPEGLRLHRAQVQIAIMTFKDYFCWNTFVENHETERK  
AWEGLEHNSVRLSRIRILLPLVEVDLDRDAFRITGL"  
BASE COUNT 127 a 165 c 155 g 149 t  
ORIGIN

Query Match 99.6%; Score 594.4; DB 9; Length 596;  
Best Local Similarity 99.8%; Pred. No. 1.3e-149;  
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGACACGCTCTTGATGAACCGGAGGAGTTCCTTACCAATTCAAAATGTCGCTGG 60  
DB |  
1 ATGACACGCTCTTGATGAACCGGAGGAGTTCCTTACCAATTCAAAATGTCGCTGG 60  
QY 61 GCTAAGGTCGCGGTGAGACCTACCTGTGTACGTAGAGAGCGGTGACAGTGGCTC 120  
DB |  
61 GCTAAGGTCGCGGTGAGACCTACCTGTGTACGTAGAGAGCGGTGACAGTGGCTC 120  
QY 121 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAACGCGTGCACAGTGGATTGCTC 180  
DB |  
121 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAACGCGTGCACAGTGGATTGCTC 180  
QY 181 TTCCTCCGCTACATCTCGGACTGGACCTAGACCCCTGGCGGTGCTACCGCGTCACTGG 240  
DB |  
181 TTCCTCCGCTACATCTCGGACTGGACCTAGACCCCTGGCGGTGCTACCGCGTCACTGG 240  
QY 241 TTCACCTCTCGGAGCCCTCTGCTACGACTGTGCGCGGCTGCTACCGCGTCACTGG 300  
DB |  
241 TTCACCTCTCGGAGCCCTCTGCTACGACTGTGCGCGGCTGCTACCGCGTCACTGG 300





BASE COUNT		126 a	164 c	155 g	151 t	/db_xref="taxon:9606"	
ORIGIN						/clone="Ramos 7"	
Query Match		99.6%	Score 594.4;	DB 9;	Length 596;	/gene="AID"	
Best Local Similarity		99.8%;	Pred. No. 1.3e-149;			/gene="AID"	
Matches 595;		Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;	/note="Integrated into Burkitt's lymphoma cell line Ramos"	
						/codon_start=1	
						/product="AID"	
						/protein_id="AA95408.1"	
						/db_xref="GI:2297230"	
BASE COUNT		127 a	165 c	155 g	149 t		
ORIGIN							
Query Match		99.6%	Score 594.4;	DB 9;	Length 596;		
Best Local Similarity		99.8%;	Pred. No. 1.3e-149;				
Matches 595;		Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;		
Qy	1	ATGCACAGCCTCTTGATGAACCGGAGAAAGTTCTTTACCAATTTCAAAATGTCGCTGG	60				
Db	1	ATGCACAGCCTCTTGATGAACCGGAGAAAGTTCTTTACCAATTTCAAAATGTCGCTGG	60				
Qy	61	GCTAAGGGTCGGGCTGAGACCTACCTGTGCTAGTGAAGAGCGGTGACAGTGGCTTACA	120				
Db	61	GCTAAGGGTCGGGCTGAGACCTACCTGTGCTAGTGAAGAGCGGTGACAGTGGCTTACA	120				
Qy	121	TCCTTTTCACTGGACTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAATTTGCTC	180				
Db	121	TCCTTTTCACTGGACTTGGTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAATTTGCTC	180				
Qy	181	TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGTCTACCGGCTACCTGG	240				
Db	181	TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGTCTACCGGCTACCTGG	240				
Qy	241	TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGCAGG	300				
Db	241	TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGCAGG	300				
Qy	301	AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	360				
Db	301	AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG	360				
Qy	361	GCTGAGCCCGAGGGCTGCGGGGCTGCACCGCGCGGTGCAATACGCATCATGACC	420				
Db	361	GCTGAGCCCGAGGGCTGCGGGGCTGCACCGCGCGGTGCAATACGCATCATGACC	420				
Qy	421	TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAAACCATCAAAAGAACTTTCAA	480				
Db	421	TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAAACCATCAAAAGAACTTTCAA	480				
Qy	481	GCCTGGGAAGGGCTGCATGAAAATTCAGTTCGTTCTCTCCAGACAGCTTCGGGCGATCCTT	540				
Db	481	GCCTGGGAAGGGCTGCATGAAAATTCAGTTCGTTCTCTCCAGACAGCTTCGGGCGATCCTT	540				
Qy	541	TTGCCCCCTGTATGAGTTGATGACTTACGAGAGCGAATTTTCGACTTTTGGGACTTTG	596				
Db	541	TTGCCCCCTGTATGAGTTGATGACTTACGAGAGCGAATTTTCGACTTTTGGGACTTTG	596				
RESULT 9							
AF529822							
LOCUS							
DEFINITION							
ACCESSION							
VERSION							
KEYWORDS							
SOURCE							
ORGANISM							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
REFERENCE							
AUTHORS							
TITLE							
JOURNAL							
FEATURES							
source							

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
AUTHORS 1 (bases 1 to 596)  
TITLE Martin, A. and Scharff, M.D.  
Somatic hypermutation of the AID transgene in B cells and non-B cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES Location/Qualifiers  
source  
1. .596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Ramos 8"  
1. .596  
/gene="AID"  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95409.1"  
/db\_xref="GI:22297232"  
/translation="MDSLNNRRKFLYQFNVRWAKGRRETYLCYVVKRRDRATSPSL  
DFGLRNKNGCHVELLFLRYISDWLDLPGRCYRVWTFWSPCYDCARHVADELRGNP  
NLSLRFTARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYFCWNTFVENHERTEK  
AWEGHNSVRLSRQLRILLPLYEVDLDAFRTGLL"  
BASE COUNT 126 a 165 c 155 g 150 t  
ORIGIN  
Query Match 99.6%; Score 594.4; DB 9; Length 596;  
Best Local Similarity 99.8%; Pred. No. 1.3e-149;  
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGCACAGCCTCTTGATGAACCGGAGGAAAGTTTCTTTTACCAATTCAAAAATGCCGCTGG 60  
DB 1 ATGCACAGCCTCTTGATGAACCGGAGGAAAGTTTCTTTTACCAATTCAAAAATGCCGCTGG 60  
QY 61 GCTAAGGGTCGGCTGAGACCTACCTGCTAGTGTAGTGAAGAGGGTGACAGTGCTTACA 120  
DB 61 GCTAAGGGTCGGCTGAGACCTACCTGCTAGTGTAGTGAAGAGGGTGACAGTGCTTACA 120  
QY 121 TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATTCGCTC 180  
DB 121 TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATTCGCTC 180  
QY 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 240  
DB 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 240  
QY 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCCGGACATGCTGGCCGACTTTCGCGAGGG 300  
DB 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCCGGACATGCTGGCCGACTTTCGCGAGGG 300  
QY 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
DB 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
QY 361 GCTGAGCCCGAGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGAC 420  
DB 361 GCTGAGCCCGAGGGCTGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGAC 420  
QY 421 TTCAAAGATTTATTTTACTGCTGGAATACCTTTTGTAGAAAACCATGAAAGAACTTTCAAA 480  
DB 421 TTCAAAGATTTATTTTACTGCTGGAATACCTTTTGTAGAAAACCATGAAAGAACTTTCAAA 480  
QY 481 GCTGAGGAGGGCTGATGAAATTCAGTTCGCTCTCTCCAGACAGCTTGGCGCATCCTT 540  
DB 481 GCTGAGGAGGGCTGATGAAATTCAGTTCGCTCTCTCCAGACAGCTTGGCGCATCCTT 540  
QY 541 TTGCCCTCTGTATGAGCTTGATGACTTACGAGACGCATTTCTGTTTGGGACTTTG 596

Db 541 TTGCCCTCTGTATGAGCTTGATGACTTACGAGACGCATTTCTGTTTGGGACTTTG 596  
RESULT 10  
AF529823  
LOCUS Homo sapiens clone Ramos 9 AID (AID) mRNA, partial cds.  
DEFINITION  
AF529823  
ACCESSION  
AF529823.1 GI:22297233  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES Location/Qualifiers  
source  
1. .596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Ramos 9"  
1. .596  
/gene="AID"  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95410.1"  
/db\_xref="GI:22297234"  
/translation="MDSLNNRRKFLYQFNVRWAKGRRETYLCYVVKRRDRATSPSL  
DFGLRNKNGCHVELLFLRYISDWLDLPGRCYRVWTFWSPCYDCARHVADELRGNP  
NLSLRFTARLYFCEDRKAEPGLRLHRAGVQIAIMTFKDYFCWNTFVENHERTEK  
AWEGHNSVRLSRQLRILLPLYEVDLDAFRTGLL"  
BASE COUNT 128 a 163 c 155 g 150 t  
ORIGIN  
Query Match 99.6%; Score 594.4; DB 9; Length 596;  
Best Local Similarity 99.8%; Pred. No. 1.3e-149;  
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGCACAGCCTCTTGATGAACCGGAGGAAAGTTTCTTTTACCAATTCAAAAATGCCGCTGG 60  
DB 1 ATGCACAGCCTCTTGATGAACCGGAGGAAAGTTTCTTTTACCAATTCAAAAATGCCGCTGG 60  
QY 61 GCTAAGGGTCGGCTGAGACCTACCTGCTAGTGTAGTGAAGAGGGTGACAGTGCTTACA 120  
DB 61 GCTAAGGGTCGGCTGAGACCTACCTGCTAGTGTAGTGAAGAGGGTGACAGTGCTTACA 120  
QY 121 TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATTCGCTC 180  
DB 121 TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAATTCGCTC 180  
QY 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 240  
DB 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 240  
QY 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCCGGACATGCTGGCCGACTTTCGCGAGGG 300  
DB 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCCGGACATGCTGGCCGACTTTCGCGAGGG 300  
QY 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
DB 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360

```
Qy 361 GCTGAGCCCGAGGGGTGCGGGCTGCACCGCGCCGGGTGCAAAATAGCCATCATGACC 420
|||||
Db 361 GCTGAGCCCGAGGGGTGCGGGCTGCACCGCGCCGGGTGCAAAATAGCCATCATGACC 420
|||||
Qy 421 TTCAAGATTTATTTTACTGCTGGAATACHTTTGTAGAAAACCATGAAGAAGCTTTCAAA 480
|||||
Db 421 TTCAAGATTTATTTTACTGCTGGAATACHTTTGTAGAAAACCATGAAGAAGCTTTCAAA 480
|||||
Qy 481 GCCTGGAAGGCTGTCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
|||||
Db 481 GCCTGGAAGGCTGTCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
|||||
Qy 541 TTGCCCTGTATGAGTTGATGACTTACGAGAGCGCATTTTCGTACTTTGGGACTTTG 596
|||||
Db 541 TTGCCCTGTATGAGTTGATGACTTACGAGAGCGCATTTTCGTACTTTGGGACTTTG 596
|||||

RESULT 11
AF529824
LOCUS Homo sapiens clone Ramos 10 AID (AID) mRNA, linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 10 AID (AID) mRNA, partial cds.
ACCESSION AF529824
VERSION AF529824.1 GI:22297235
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 10"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95411.1"
/db_xref="GI:22297236"
/translation="MDSLMNRKFLYQFKNVRWAKGRREYVLCYVVKRDSATPSL
DFGYLRNKGHWELFLRYISDWLDPCRYVFTWSWPCYDCARHVPFLRGNP
NLSLRIFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCYCNWTFVENHERTFK
AMEGLHENSVRSLRLLPLLYEVDLDRAPFTLGL"
BASE COUNT 127 a 163 c 156 g 150 t
ORIGIN

Query Match 99.6%; Score 594.4; DB 9; Length 596;
Best Local Similarity 99.8%; Pred. No. 1.3e-149;
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ATGCACAGCTCTTGATGAACCGGAGGAGTTCTTTTACCAATTCAAAATGTCCTGG 60
|||||
Db 1 ATGCACAGCTCTTGATGAACCGGAGGAGTTCTTTTACCAATTCAAAATGTCCTGG 60
|||||
Qy 61 GCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGTACA 120
|||||
Db 61 GCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGGCGTGACAGTGTACA 120
|||||
Qy 121 TCCTTTTACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCCTC 180
|||||
```

```
Db 121 TCCTTTTACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCCTC 180
|||||
Qy 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCTCACCTGG 240
|||||
Db 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCTCACCTGG 240
|||||
Qy 241 TTCACCTCTCGGAGCCCTCTGCTACGACTGTGCCCGACATGTGCCGAGCTTTCTCGGAGGG 300
|||||
Db 241 TTCACCTCTCGGAGCCCTCTGCTACGACTGTGCCCGACATGTGCCGAGCTTTCTCGGAGGG 300
|||||
Qy 301 AACCCCAACCTCAGTCTGAGGATCTTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
|||||
Db 301 AACCCCAACCTCAGTCTGAGGATCTTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
|||||
Qy 361 GCTGAGCCCGAGGGGTGCGGGCTGCACCGCGCGGTGCANATAGCCATCATGACC 420
|||||
Db 361 GCTGAGCCCGAGGGGTGCGGGCTGCACCGCGCGGTGCANATAGCCATCATGACC 420
|||||
Qy 421 TTCAAGATTTATTTTACTGCTGGAATACHTTTGTAGAAAACCATGAAGAAGCTTTCAAA 480
|||||
Db 421 TTCAAGATTTATTTTACTGCTGGAATACHTTTGTAGAAAACCATGAAGAAGCTTTCAAA 480
|||||
Qy 481 GCCTGGAAGGCTGTCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
|||||
Db 481 GCCTGGAAGGCTGTCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
|||||
Qy 541 TTGCCCTGTATGAGTTGATGACTTACGAGAGCGCATTTTCGTACTTTGGGACTTTG 596
|||||
Db 541 TTGCCCTGTATGAGTTGATGACTTACGAGAGCGCATTTTCGTACTTTGGGACTTTG 596
|||||

RESULT 12
AF529825
LOCUS Homo sapiens clone Ramos 11 AID (AID) mRNA, linear PRI 19-AUG-2002
DEFINITION Homo sapiens clone Ramos 11 AID (AID) mRNA, partial cds.
ACCESSION AF529825
VERSION AF529825.1 GI:22297237
KEYWORDS human.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Ramos 11"
1..>596
/gene="AID"
1..>596
/gene="AID"
/notes="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95412.1"
/db_xref="GI:22297238"
/translation="MDSLMNRKFLYQFKNVRWAKGRREYVLCYVVKRDSATPSL
DFGYLRNKGHWELFLRYISDWLDPCRYVFTWSWPCYDCARHVPFLRGNP
NLSLRIFTARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCYCNWTFVENHERTFK
AMEGLHENSVRSLRLLPLLYEVDLDRAPFTLGL"
BASE COUNT 126 a 164 c 156 g 150 t
ORIGIN
```

Query Match		99.6%; Score 594.4; DB 9; Length 596;	CDS	/gene="AID"			
Best Local Similarity		99.8%; Pred. No. 1.3e-149;		1. .>596			
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				/gene="AID"			
				/note="integrated into Burkitt's lymphoma cell line Ramos"			
				/codon_start=1			
				/product="AID"			
				/protein_id="AAM95413.1"			
				/db_xref="GI:22297240"			
				/translation="MDSLIMNRKFLYQFKNVRWAKGRRETYLCVYVKKRDSATSPSL			
				DFGYLRNKGCHVELFLRYISDWLDPRGCRVYTFWTSWSPCYDCARHVAQFLRGNP			
				NLSLRIFEARLYFCEDRRKAPGELRLRHRAGVOIAIMTKDYFYCHWNTFVENHERTEFK			
				AWEGUHSNVRSLRQLRILLPLLYEVDLDRADFRTGL"			
BASE COUNT		127 a 163 c 155 g 151 t					
ORIGIN							
Query Match		99.6%; Score 594.4; DB 9; Length 596;					
Best Local Similarity		99.8%; Pred. No. 1.3e-149;					
Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;							
QY	1	ATGGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGG	60				
Db	1	ATGGACAGCCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTCAAAAATGTCGCTGG	60				
QY	61	GCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA	120				
Db	61	GCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA	120				
QY	121	TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC	180				
Db	121	TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC	180				
QY	181	TTCTCCCGTACATCTCGGACTTGGGACCTAGACCTGGCCGCTGTACCGCTGCACCTGG	240				
Db	181	TTCTCCCGTACATCTCGGACTTGGGACCTAGACCTGGCCGCTGTACCGCTGCACCTGG	240				
QY	241	TTCAAGGTCGAGCGCTGCTACGATCTGCGCAATAAGAACGGCTGCCACGTGGAATTGCTC	300				
Db	241	TTCAAGGTCGAGCGCTGCTACGATCTGCGCAATAAGAACGGCTGCCACGTGGAATTGCTC	300				
QY	301	RACCCCAACCTCAGTCTCAGGATCTTTCACCGCGCGCCTCTACTTCTGTGAGGACCGCAAG	360				
Db	301	RACCCCAACCTCAGTCTCAGGATCTTTCACCGCGCGCCTCTACTTCTGTGAGGACCGCAAG	360				
QY	361	GCTGAGCCGAGGGCTGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC	420				
Db	361	GCTGAGCCGAGGGCTGCGGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC	420				
QY	421	TTCAAGATTTATTTTACTGCTGGAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTT	480				
Db	421	TTCAAGATTTATTTTACTGCTGGAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTT	480				
QY	481	GCCTGGGAGGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTT	540				
Db	481	GCCTGGGAGGGCTGCATGAAATTCAGTTCGCTCTCCAGACAGCTTCGGCGCATCTT	540				
QY	541	TTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG	596				
Db	541	TTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTG	596				
RESULT 13							
AF529826							
LOCUS		596 bp mRNA linear PRI 19-AUG-2002					
DEFINITION		Homo sapiens clone Ramos 12 AID (AID) mRNA, partial cds.					
ACCESSION		AF529826					
VERSION		AF529826.1 GI:22297239					
KEYWORDS		human.					
SOURCE		Homo sapiens					
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
REFERENCE		1 (bases 1 to 596)					
AUTHORS		Martin,A. and Scharff,M.D.					
TITLE		Somatic hypermutation of the AID transgene in B cells and non-B cells					
JOURNAL		Unpublished					
REFERENCE		2 (bases 1 to 596)					
AUTHORS		Martin,A. and Scharff,M.D.					
TITLE		Direct Submission					
JOURNAL		Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA					
FEATURES		Location/Qualifiers					
source		1. .596					
		/organism="Homo sapiens"					
		/db_xref="taxon:9606"					
		/clone="Ramos 12"					
		1. .>596					
gene							

TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 596)

AUTHORS Martin,A. and Scharff,M.D.

TITLE Direct Submission

JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES

Location/Qualifiers

1..596

Source

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="Ramos 13"

1..>596

/gene="AID"

1..>596

/gene="AID"

/note="Integrated into Burkitt's lymphoma cell line Ramos"

/codon\_start=1

/product="AID"

/protein\_id="AAM95414.1"

/db\_xref="GI:22297242"

translation="MDSLIMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKGCHVELLFURYISDMDLDPGRCYRVTWFTSWSPCYDCARHVDPLRGNP NLSLRIFARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHERTEK AWEGUHSNLSRLRLPLLEVDLDRDAPRTLGL"

BASE COUNT 127 a 163 c 155 g 151 t

ORIGIN

Query Match 99.6%; Score 594.4; DB 9; Length 596;

Best Local Similarity 99.8%; Pred. No. 1.3e-149;

Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGACACCCCTTCATGACACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGGCTGG 60

Db 1 ATGGACACCCCTTCATGACACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGGCTGG 60

QY 61 GCTAAGGTCGGCGTGAGACCTACCTGTGTACCTAGTGTGAGAGGCGTGACAGTGTCTACA 120

Db 61 GCTAAGGTCGGCGTGAGACCTACCTGTGTACCTAGTGTGAGAGGCGTGACAGTGTCTACA 120

QY 121 TCCTTTTACCTGGACTTTGGTTATCTTCGCAATAAGAACGCGTGCACAGTGGAAATTCCTC 180

Db 121 TCCTTTTACCTGGACTTTGGTTATCTTCGCAATAAGAACGCGTGCACAGTGGAAATTCCTC 180

QY 181 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGTGCTACCTAGTGAAGAGCGGTGACAGTGTCTA 240

Db 181 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGTGCTACCTAGTGAAGAGCGGTGACAGTGTCTA 240

QY 241 TTCAAGAGATATTTTACTGCTGGAATATCTTTGTAGAAACCATCAAGAACTTTCAAA 480

Db 241 TTCAAGAGATATTTTACTGCTGGAATATCTTTGTAGAAACCATCAAGAACTTTCAAA 480

QY 481 GCCTGGGAAGGCGTCGATGAAATTCAGTTCTGCTCTCCAGACACCTTCGGGCGCATCTTT 540

Db 481 GCCTGGGAAGGCGTCGATGAAATTCAGTTCTGCTCTCCAGACACCTTCGGGCGCATCTTT 540

QY 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTTCTGACTTTGGGACTTTG 596

Db 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTTCTGACTTTGGGACTTTG 596

RESULT 15

AF529829

LOCUS

DEFINITION Mus musculus clone 2 transgenic Homo sapiens AID (AID) mRNA, SYN 19-AUG-2002 complete cds.

ACCESSION AF529829

VERSION AF529829.1

KEYWORDS GI:22297245

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 596)

Martin,A. and Scharff,M.D.

Somatic hypermutation of the AID transgene in B and non-B cells

Unpublished

2 (bases 1 to 596)

Martin,A. and Scharff,M.D.

Direct Submission

Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA

FEATURES

Location/Qualifiers

1..596

Source

/organism="Mus musculus"

/db\_xref="taxon:10090"

/clone="2"

/cell\_type="hybridoma pl-5"

/transgenic

1..596

/organism="Homo sapiens"

/db\_xref="taxon:9606"

1..>596

/gene="AID"

1..>596

/gene="AID"

/codon\_start=1

/product="AID"

/protein\_id="AAM95416.1"

/db\_xref="GI:22297246"

translation="NDRLLMNRKFLYOFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKGCHVELLFURYISDMDLDPGRCYRVTWFTSWSPCYDCARHVDPLRGNP NLSLRIFARLYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHERTEK AWEGUHSNLSRLRLPLLEVDLDRDAPRTLGL"

BASE COUNT 128 a 163 c 155 g 150 t

ORIGIN

Query Match 99.6%; Score 594.4; DB 12; Length 596;

Best Local Similarity 99.8%; Pred. No. 1.3e-149;

Matches 595; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGGACACCCCTTCATGACACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGGCTGG 60

Db 1 ATGGACACACCTTCATGACACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGGCTGG 60

QY 61 GCTAAGGTCGGCGTGAGACCTACCTGTGCTACCTAGTGAAGAGCGGTGACAGTGTCTACA 120

Db 61 GCTAAGGTCGGCGTGAGACCTACCTGTGCTACCTAGTGAAGAGCGGTGACAGTGTCTACA 120

QY 121 TCCTTTTACCTGGACTTTGGTTATCTTCGCAATAAGAACGCGTGCACAGTGGAAATTCCTC 180

Db 121 TCCTTTTACCTGGACTTTGGTTATCTTCGCAATAAGAACGCGTGCACAGTGGAAATTCCTC 180

QY 181 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGTGCTACCTAGTGAAGAGCGGTGACAGTGTCTA 240

Db 181 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGTGCTACCTAGTGAAGAGCGGTGACAGTGTCTA 240

QY 241 TTCAAGAGATATTTTACTGCTGGAATATCTTTGTAGAAACCATCAAGAACTTTCAAA 300

Db 241 TTCAAGAGATATTTTACTGCTGGAATATCTTTGTAGAAACCATCAAGAACTTTCAAA 300

QY 301 AACCCCAACCTCAGTGTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360

Db 301 AACCCCAACCTCAGTGTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360

QY 361 GCTGAGCCGAGGGCGTGCGGGCGTGACCGCGCGGGGTGCAATATGCCATCATGACC 420

Db 361 GCTGAGCCGAGGGCGTGCGGGCGTGACCGCGCGGGGTGCAATATGCCATCATGACC 420

QY 421 TTCAAGAGATATTTTACTGCTGGAATATCTTTGTAGAAACCATCAAGAACTTTCAAA 480

Db 421 TTCAAGAGATATTTTACTGCTGGAATATCTTTGTAGAAACCATCAAGAACTTTCAAA 480

QY 481 GCCTGGGAAGGCGTCGATGAAATTCAGTTCTGCTCTCCAGACACCTTCGGGCGCATCTTT 540

Db 481 GCCTGGGAAGGCGTCGATGAAATTCAGTTCTGCTCTCCAGACACCTTCGGGCGCATCTTT 540

QY 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTTCTGACTTTGGGACTTTG 596

Db 541 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTTCTGACTTTGGGACTTTG 596

Db	301	AACCCCAACCTCAGTCTGAGGATCTTACCGGGCGCCTCTACTTCTGTGAGGACCGCAAG	360
Qy	361	GCTGAGCCCGAGGGGCTGCGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC	420
Db	361		420
Qy	421	TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTTCAA	480
Db	421		480
Qy	481	GCCTGGGAAGGCTGCATGAAAATTCAGTTCGCTCTCCAGACAGCTTCGGGCGATCCTT	540
Db	481		540
Qy	541	TTGCCCTGTATGAGGTTGATGACTTACGAGACGATTTCTACTTTGGGACTTTG	596
Db	541		596

Search completed: June 19, 2003, 07:46:31  
Job time : 1525.43 secs





Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	597	100.0		2818	21	AAC55312	Mouse activation-i
2	475.4	79.6		2440	21	AAC55307	Mouse activation-i
3	273.4	45.8		6564	21	AAC55314	Human activation-i
4	273.4	45.8		11204	21	AAC55339	Human activation-i
5	271	45.4		271	21	AAC55317	Human activation-i
6	148	24.8		148	21	AAC55316	Human activation-i
7	118.4	19.8		1534	20	AAZ20836	Polynucleotide seq
8	118.4	19.8		1534	22	AAS59293	Human CDNA encodin
9	118.4	19.8		1534	24	ABA30962	Human polynucleoti

PF 28-MAR-2000; 2000WO-JP01918.  
 XX 29-MAR-1999; 99JP-0087192.  
 PR 24-JUN-1999; 99JP-0178999.  
 PR 27-DEC-1999; 99JP-0371382.  
 XX  
 PA (NISB ) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 PA Honjo T, Muramatsu M;  
 PI WPI; 2000-611715/58.  
 XX P-PSDB; AAB24198.  
 DR  
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 XX  
 PS Claim 3; Page 135-139; 174pp; Japanese.  
 XX  
 CC The present sequence encodes human activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and  
 CC dermatological activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
 CC immunodeficiency disorder, MHC (major histocompatibility class) class  
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
 CC encoding AID may be used for gene therapy and the antibodies to the AID  
 CC protein may be used for diagnosis and treatment of these disorders.  
 XX  
 SQ Sequence 2818 BP; 868 A; 548 C; 626 G; 776 T; 0 other;  
 Query Match 100.0%; Score 597; DB 21; Length 2818;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-167;  
 Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 ATGACAGCCTCTGTATGAACCGGAGGAGTCTTTTACCAATTCAAAATGTCGCTGG 60  
 DB |  
 QY 80 ATGACAGCCTCTGTATGAACCGGAGGAGTCTTTTACCAATTCAAAATGTCGCTGG 139  
 DB |  
 QY 61 GCTAAGGTCGCGGTGAGACCTACCTGCTAGTGTAGTGAAGAGCGGTGACATGTGTACA 120  
 DB |  
 QY 140 GCTAAGGTCGCGGTGAGACCTACCTGCTAGTGTAGTGAAGAGCGGTGACATGTGTACA 199  
 QY 121 TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAAATGCTC 180  
 DB |  
 QY 200 TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGAAATGCTC 259  
 DB |  
 QY 181 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGTACCGCTCACTGG 240  
 DB |  
 QY 260 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGTACCGCTCACTGG 319  
 DB |  
 QY 241 TTCACCTCTCGGACCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAGGG 300  
 DB |  
 QY 320 TTCACCTCTCGGACCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAGGG 379  
 QY 301 AACCCCAACCTCAGTGTAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAG 360  
 DB |  
 QY 380 AACCCCAACCTCAGTGTAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAAG 439  
 QY 361 GCTGAGCCGAGGGGCTGCGGGGCTGTACCGCGCGCGGGTGCATAATAGCCATCATGACC 420  
 DB |  
 QY 440 GCTGAGCCGAGGGGCTGCGGGGCTGTACCGCGCGCGGGTGCATAATAGCCATCATGACC 499  
 DB |  
 QY 421 TTCAAGATATTATTTTACTGCTGGAATATCTTTGTAGAAAACCATCAAGAAGACTTCAA 480  
 DB |

DB 500 TTCAAGATATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 559  
 QY 481 GCCTGGGAAGGCTGCATCAAAATTCAGTTCTGCTCTCCAGACAGCTTCGGGCGCATCTT 540  
 DB |  
 QY 560 GCCTGGGAAGGCTGCATCAAAATTCAGTTCTGCTCTCCAGACAGCTTCGGGCGCATCTT 619  
 DB |  
 QY 541 TTGCCCTCTGATGAGTTGATGACATTACGAGACGCAATTCGTACTTTGGGACTTTGA 597  
 DB |  
 QY 620 TTGCCCTCTGATGAGTTGATGACATTACGAGACGCAATTCGTACTTTGGGACTTTGA 676  
 DB |  
 RESULT 2  
 AAC55307  
 ID AAC55307 standard; cDNA; 2440 BP.  
 XX  
 AC AAC55307;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX  
 DE Mouse activation-induced cytidine deaminase encoding cDNA SEQ ID NO:1.  
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; antiallergic;  
 KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease;  
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ss.  
 XX  
 OS Mus musculus.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 93..589  
 FT /\*tag= a  
 FT /product= "activation-induced cytidine deaminase"  
 XX  
 PN WO200058480-A1.  
 XX  
 XX 05-OCT-2000.  
 XX  
 PD 28-MAR-2000; 2000WO-JP01918.  
 PF  
 XX 29-MAR-1999; 99JP-0087192.  
 PR  
 PR 24-JUN-1999; 99JP-0178999.  
 PR  
 PR 27-DEC-1999; 99JP-0371382.  
 XX  
 PA (NISB ) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 XX  
 XX Honjo T, Muramatsu M;  
 PI WPI; 2000-611715/58.  
 XX P-PSDB; AAB24197.  
 DR  
 DR Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 XX  
 PS Claim 3; Page 126-130; 174pp; Japanese.  
 XX  
 CC The present sequence encodes mouse activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and  
 CC dermatological activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen

CC	disease, DiGeorge disease, ataxia telangiectasia, common variable
CC	immunodeficiency disorder, MHC (major histocompatibility class)
CC	II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated
CC	IgE disorder, and IgG subclass selection disorder. The DNA sequences
CC	encoding AID may be used for gene therapy and the antibodies to the AID
CC	protein may be used for diagnosis and treatment of these disorders.
XX	
SQ	Sequence 2440 BP; 706 A; 546 C; 551 G; 636 T; 1 other;
Query Match 79.6%; Score 475.4; DB 21; Length 2440;	
Best Local Similarity 87.3%; Pred. No. 4.5e-131;	
Matches 521; Conservative 0; Mismatches 76; Indels 0; Gaps 0;	
QY	1 ATGCACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGTGG 60
DB	
DB	93 ATGCACAGCCTCTTGATGAAGCAAAAGAAGTTCTTTACCAATTCAAAAATGTCGGTGG 152
QY	61 GCTAAGGGTCGGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGCTGACAGTGTCTACA 120
DB	
DB	153 GCCAAGGGACGGATGAGACTACCTGTGCTACGTGTGTTGAAGAGGAGATGATGCCACC 212
QY	121 TCCTTTTCACCTGGACTTTGGTTACTTCGCAATTAAGAACGGCTGCCAGCTGGAATTGGCT 180
DB	
DB	213 TCCTGCTCACTGGACTTCGGCCACCTTCGCAACAAGCTTGCTGCCAGCTGGAATGTGTG 272
QY	181 TTCCTCGCTACATCTCGGACTGGGACCTAGACCTTGCCGCTCTACCGGCTCACCTGG 240
DB	
DB	273 TTCCTACGCTACATCTCAGACTGGGACCTGGACCCGGCGCGTGTACCGCTCACCTGG 332
QY	241 TTCACCTCTCGAGCCCTCTGTCAGCACTGTGCCGACATGTGGCGGACCTTTCTCGGAGGG 300
DB	
DB	333 TTCACCTCTCGAGCCCGTCTATGACTGTGCCGCGCACGTTGGCTGAGTTTCTGAGATGG 392
QY	301 AACCCCAACCTCAGCTCGAGGATCTTCACGCGCGCCTCTACTTCTGTGAGAGCCGCAAG 360
DB	
DB	393 AACCTAACCTCAGCTGAGGATTTTACCGCGCGCCTCTACTTCTGTGAAGACCGCAAG 452
QY	361 GCTGAGCCCGAGGGGCTCGCGCGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
DB	
DB	453 GCTGAGCCTGAGGGGCTCGGAGACTGCACCGCGCTGGGGTCCAGATCGGCGCATGACC 512
QY	421 TTCAAAGATATTTTTACTGCTGGAATACTTTTGTAGNAACCATGAAGAAGAACTTCAAA 480
DB	
DB	513 TTCAAAGACTATTTTTACTGCTGGAATACATTTGTAGAAAAATCGTGAAGAAGAACTTCAAA 572
QY	481 GCCTGGGAAGGGCTGCAATAAAATTCAGTTGCTCTCCACACAGCTTCGGCGCATCCTT 540
DB	
DB	573 GCCTGGGAAGGGCTACATGAATAATCTCTCGGGCTAACCCACACAACTTCGGCGCATCCTT 632
QY	541 TTGCCCCGTATGAGGTGTGATGACTTACGAGACGCAATTTCTGCTACTTTGGGACTTTGA 597
DB	
DB	633 TTGCCCCGTACGAAGTCGATGACTTCGGAGATGCATTTTCGTATGTTGGGATTTTGA 689
RESULT 3	
AAC55314	
ID	AAC55314 standard; DNA: 6564 BP.
XX	
AC	AAC55314;
XX	
XX	
DT	05-FEB-2001 (first entry)
DE	
XX	Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:10.
XX	
KW	Activation-induced cytidine deaminase; AID; cytidine deaminase;
KW	immune related disease; allergy; allergic disease; anti-allergic;
KW	antianemic; antistaphylococcal; ophthalmological; anti-HIV; dermatological;
KW	gene therapy; B cell associated immune system disorder; food allergy;
KW	immunodeficiency disease; immunoglobulin A deficiency disease; asthma;
KW	IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;
KW	drug allergy; allergic rhinitis; rosacea disease; DiGeorge disease; AIDS;
KW	ataxia telangiectasia; common variable immunodeficiency disorder;
KW	major histocompatibility class II deficiency disease;

AAC55339  
 ID AAC55339 standard; DNA; 11204 BP.  
 XX  
 AC AAC55339;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX  
 DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.  
 XX  
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; antiallergic;  
 KW antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease;  
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.  
 OS Homo sapiens.  
 XX  
 XX  
 PN WO200058480-A1.  
 XX  
 XX  
 PD 05-OCT-2000.  
 XX  
 XX  
 PF 28-MAR-2000; 2000WO-JP01918.  
 XX  
 PR 29-MAR-1999; 99JP-0087192.  
 PR 24-JUN-1999; 99JP-0178999.  
 PR 27-DEC-1999; 99JP-0371382.  
 XX  
 XX  
 PA (NIBS) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 XX  
 PI Honjo T, Muramatsu M;  
 PI  
 XX  
 DR WPI; 2000-611715/58.  
 XX  
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 XX  
 PS Claim 17; Page 163-170; 174pp; Japanese.  
 XX  
 CC The present invention describes an activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
 CC antianaemic, antiasthmatic, ophthalmological, anti-HIV and  
 CC dermatological activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
 CC immunodeficiency disorder, MHC (major histocompatibility class) class  
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
 CC Ig disorder, and IgG subclass selection disorder. The DNA sequences  
 CC encoding AID may be used for gene therapy and the antibodies to the AID  
 CC protein may be used for diagnosis and treatment of these disorders. The  
 CC present sequence represents a genomic DNA sequence of human AID.  
 XX  
 SQ Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 other;  
 Query Match 45.8%; Score 273.4; DB 21; Length 11204;  
 Best Local Similarity 99.6%; Pred. No. 1.3e-70;  
 Matches 274; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 155 AGACGGCTGCCACGTGGAATTCCTTCCTCCGCTACATCTCGGACGTGGACCTAGACC 214  
 DB 7805 AGACGGCTGCCACGTGGAATTCCTTCCTCCGCTACATCTCGGACGTGGACCTAGACC 7864

QY 215 CTGGCGCGCTGCTACCGCGTCACTGGTTCACTCTCGAGCCCTGCTACGACTGTGCC 274  
 DB 7865 CTGGCGCGTGTCTACCGCGTCACTGGTTCACTCTCGAGCCCTGCTACGACTGTGCC 7924  
 QY 275 GACATGTGGCGGACTTCTCTGGCGAGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGC 334  
 DB 7925 GACATGTGGCGGACTTCTCTGGCGAGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGC 7984  
 QY 335 GCCTTACTCTGTGTGAGGACCGCAAGGCTGAGCCCGAGGGCTGGCGGGTGCACACCGC 394  
 DB 7985 GCCTTACTCTGTGTGAGGACCGCAAGGCTGAGCCCGAGGGCTGGCGGGTGCACACCGC 8044  
 QY 395 CGGGGTGCAATACCCATCATGACCTTCAAGAT 429  
 DB 8045 CGGGGTGCAATACCCATCATGACCTTCAAGAT 8079  
 RESULT 5  
 AAC55317  
 ID AAC55317 standard; DNA; 271 BP.  
 XX  
 AC AAC55317;  
 XX  
 DT 05-FEB-2001 (first entry)  
 XX  
 XX  
 DE Human activation-induced cytidine deaminase exon 3 SEQ ID NO:13.  
 XX  
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; antiallergic;  
 KW antianaemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease;  
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.  
 OS Homo sapiens.  
 XX  
 XX  
 PN WO200058480-A1.  
 XX  
 XX  
 PD 05-OCT-2000.  
 XX  
 XX  
 PF 28-MAR-2000; 2000WO-JP01918.  
 XX  
 PR 29-MAR-1999; 99JP-0087192.  
 PR 24-JUN-1999; 99JP-0178999.  
 PR 27-DEC-1999; 99JP-0371382.  
 XX  
 XX  
 PA (NIBS) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 XX  
 PI Honjo T, Muramatsu M;  
 PI  
 XX  
 DR WPI; 2000-611715/58.  
 XX  
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 XX  
 PS Claim 18; Page 151; 174pp; Japanese.  
 XX  
 CC The present invention describes an activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
 CC antianaemic, antiasthmatic, ophthalmological, anti-HIV and  
 CC dermatological activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen

CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class)  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders. The  
CC present sequence represents the exon 3 genomic DNA sequence of human AID.  
XX  
SQ Sequence 271 BP; 47 A; 95 C; 76 G; 53 T; 0 other;

Query Match 45.4%; Score 271; DB 21; Length 271;  
Best Local Similarity 100.0%; Pred. No. 1.2e-70;  
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 157 AACGGCTGCCAGTGGAAATGCTCTTCCTCGCTACATCTCGGACTGGGACCTAGACCT 216  
DB 1 AACGGCTGCCAGTGGAAATGCTCTTCCTCGCTACATCTCGGACTGGGACCTAGACCT 60  
QY 217 GGCCGCTGCTACCGGCTCACCTGGTTACCTCTCTGGAGCCCTGCTACGACTGTGCCGA 276  
DB 61 GGCCGCTGCTACCGGCTCACCTGGTTACCTCTCTGGAGCCCTGCTACGACTGTGCCGA 120  
QY 277 CATGTGGCGGACTTCTCGGAGGNAACCCCAACCTCAGTCTGAGATCTTACCGCGGC 336  
DB 121 CATGTGGCGGACTTCTCGGAGGNAACCCCAACCTCAGTCTGAGATCTTACCGCGGC 180  
QY 337 CTCTACTTCTGTGAGACCGCAAGGCTGAGCCGCGGCTGCGCGGCTGCACCGGCC 396  
DB 181 CTCTACTTCTGTGAGACCGCAAGGCTGAGCCGCGGCTGCGCGGCTGCACCGGCC 240  
QY 397 GGGGTGCAATAGCCATCATGACCTTCAAAG 427  
DB 241 GGGGTGCAATAGCCATCATGACCTTCAAAG 271

RESULT 6  
AAC55316  
ID AAC55316 standard; DNA; 148 BP.  
XX  
AC AAC55316;  
DT 05-FEB-2001 (first entry)  
XX  
DE Human activation-induced cytidine deaminase exon 2 SEQ ID NO:12.  
KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; antiallergic;  
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200058480-A1.  
XX  
PD 05-OCT-2000.  
XX  
PF 28-MAR-2000; 2000WO-JP01918.  
XX  
PR 29-MAR-1999; 99JP-0087192.  
PR 24-JUN-1999; 99JP-0178999.  
PR 27-DEC-1999; 99JP-0371382.  
XX  
PA (NISR) JAPAN TOBACCO INC.  
PA (HONJ) HONJO T.  
XX  
PI Honjo T, Muramatsu M;

DR WPI; 2000-611715/58.  
XX  
PT Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
XX  
PS Claim 18; Page 150; 174pp; Japanese.  
XX

The present invention describes an activation-induced cytidine deaminase  
(AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
has cytidine activity similar to APOBEC-1. AID has antiallergic,  
CC antianemic, antiasthmatic, ophthalmological, anti-HIV and  
CC dermatological activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class) class  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders. The  
CC present sequence represents the exon 2 genomic DNA sequence of human AID.  
XX  
SQ Sequence 148 BP; 35 A; 33 C; 38 G; 42 T; 0 other;

Query Match 24.8%; Score 148; DB 21; Length 148;  
Best Local Similarity 100.0%; Pred. No. 4e-34;  
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 9 CCTCTTGATGAACCGGAGGAAGTTCTTTTACCAATTCAAAAATGTCGCTGGCTAAGGG 68  
DB 1 CCTCTTGATGAACCGGAGGAAGTTCTTTTACCAATTCAAAAATGTCGCTGGCTAAGGG 60  
QY 69 TCGCGGTGAGACCTACCTGTGCTAGTAGTGAAGAGCGCTGACAGTGTACATCTTTTC 128  
DB 61 TCGCGGTGAGACCTACCTGTGCTAGTAGTGAAGAGCGCTGACAGTGTACATCTTTTC 120  
QY 129 ACTGGACATTGGTTATCTTCGCAATAAG 156  
DB 121 ACTGGACATTGGTTATCTTCGCAATAAG 148

RESULT 7  
AAZ20856  
ID AAZ20856 standard; cDNA; 1534 BP.  
XX  
AC AAZ20856;  
DT 09-DEC-1999 (first entry)  
XX  
DE Polynucleotide sequence of the lp547\_4 clone.  
KW secreted protein; cDNA library; clone; transmembrane protein;  
KW signal sequence cloning; hybridization cloning; gene therapy;  
KW receptor; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key  
FT CDS 51..1205  
FT /tag= a  
FT /product= lp547\_4  
FT sig\_peptide 882..980  
FT /tag= b  
FT mat\_peptide 981..1205  
FT /tag= c  
XX  
PN WO9942470-A1.  
XX  
PD 26-AUG-1999.



CC contraceptive based on the ability of inhibins to decrease fertility in  
CC female mammals and decrease spermatogenesis in male mammals. The  
CC proteins and nucleic acids are also useful as food supplements. The  
CC present sequence encodes a secreted protein of the invention.  
XX  
SQ Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 other;

Query Match 19.8%; Score 118.4; DB 22; Length 1534;  
Best Local Similarity 58.9%; Pred. No. 7.7e-25;  
Matches 225; Conservative 0; Mismatches 151; Indels 6; Gaps 1;

QY 158 ACGGTGCCAGTGGAAATGCTCTTCCTCCGCTACATCTCGAGTGGACCTAGACCTGTG 217  
| ||| |||| | | |||| | | || |||| | | || |||| | | || |||| | | || |||| |  
Db 811 AAGCGCGCATGCAGAGCTGTCTCTCTGGACGTGATTCCTTTTGGGAAGCTGGACCTGG 870  
| |||| | | |||| | | |||| | | || |||| | | || |||| | | || |||| | | || |||| |  
QY 218 GCGCTGTACCGGTTCACCTGTTTCACCTCCTCGAGCCCTGCTACGACTGTGCCGAC 277  
| |||| | | |||| | | |||| | | |||| | | || |||| | | || |||| | | || |||| |  
Db 871 ACCAGGACTCAGGGTTACCTGCTTCACCTCCTCGAGCCCTGCTACGCTGTGCCGAGG 930  
| |||| | | |||| | | |||| | | |||| | | || |||| | | || |||| | | || |||| |  
QY 278 ATGTGGCGACTTTCTTGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCC 337  
| |||| | | |||| | | |||| | | |||| | | || |||| | | || |||| | | || |||| |  
Db 931 AAATGGCTAAATTCATTTCAAAAACAAACACGCTGAGCCTGTGCATCTTCACTGCCGCA 990  
| |||| | | |||| | | |||| | | |||| | | || |||| | | || |||| | | || |||| |  
QY 338 TCTACTTCTGTGAGACCGCAAGGCTGAGCCGAGGGGCTCGGGGGCTGCACCGCGCG 397  
| |||| | | |||| | | |||| | | |||| | | || |||| | | || |||| | | || |||| |  
Db 991 TCTA-----TGATGATCAAGGAAGATGTCAGGAGGGGCTGCGCACCTCGCGAGGCTG 1044  
| |||| | | |||| | | |||| | | |||| | | || |||| | | || |||| | | || |||| |  
QY 398 GGGTCAATAGCCATCATGACCTTCAAGATTTATTTTACTGCTGGGAATCTTTGTAG 457  
| |||| | | |||| | | |||| | | |||| | | || |||| | | || |||| | | || |||| |  
Db 1045 GGGCAAAATTTCAATATGACATACAGTGAATTTAAGCACTGCTGGGACACCTTTGTGG 1104  
| |||| | | |||| | | |||| | | |||| | | || |||| | | || |||| | | || |||| |  
QY 458 AAAACCATGAAGAACTTTCAAAGCTTGGGAAGGGCTGCATGAATAATTCAGTCTGCTCT 517  
| |||| | | |||| | | |||| | | |||| | | || |||| | | || |||| | | || |||| |  
Db 1105 ACCACGAGGATGTCCTTCCAGCCTGGGATGGACTAGATGAGCACAGCCCAAGACCTGA 1164  
| |||| | | |||| | | |||| | | |||| | | || |||| | | || |||| | | || |||| |  
QY 518 CCAGACAGCTTCGGGGCATCT 539  
| |||| | | |||| | | |||| | | |||| | | || |||| | | || |||| | | || |||| |  
Db 1165 GTGGAGGCTGCGGGCCATTCT 1186  
| |||| | | |||| | | |||| | | |||| | | || |||| | | || |||| | | || |||| |

## RESULT 9

ABA90962

ID ABA90962 standard; cDNA; 1534 BP.

XX ABA90962;

AC ABA90962;

DT 14-FEB-2002 (first entry)

XX Human polynucleotide SEQ ID NO 173.

KW Human; clone bd306-7; ATCC number 98599; gene therapy;  
KW immune disorder; bacterial infection; fungal infection; cancer; tumour;  
KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;  
KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;  
KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnerability;  
KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;  
KW Crohn's disease; cystostatic; anti-inflammatory; immunomodulatory;  
KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory; ss.

XX Homo sapiens.

OS US2001039335-A1.

XX PN 08-NOV-2001.

XX PD 04-DEC-2000; 2000US-0729674.

XX PF 10-APR-1997; 97US-126425P.

XX PR 04-DEC-1997; 97US-067454P.

XX PR 20-DEC-1997; 97US-068379P.

XX PR 02-JAN-1998; 98US-070346P.

XX PR 07-JAN-1998; 98US-070643P.

PR 08-JAN-1998; 98US-070755P.  
PR 13-JAN-1998; 98US-071304P.  
PR 22-JAN-1998; 98US-072134P.  
PR 30-JAN-1998; 98US-073095P.  
PR 18-FEB-1998; 98US-075038P.  
PR 30-MAR-2000; 2000US-0539330.  
PR 23-NOV-1998; 98US-0197886.  
XX

PA (JACO/) JACOBS K.  
PA (MCCO/) MCCOY J M.  
PA (LAVA/) LAVALLIE E R.  
PA (COLL/) COLLINS-RACIE L A.  
PA (EVAN/) EVANS C.  
PA (MERB/) MERBERG D.  
PA (TREA/) TREACY M.  
PA (AGOS/) AGOSTINO M J.  
PA (STEI/) STEININGER R J.  
PA (SPAU/) SPAULDING V.  
PA (WONG/) WONG G G.  
PA (CLAR/) CLARK H.  
PA (FECH/) FECHTEL K.

PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;  
PI Wong GG, Clark H, Fechtel K;

XX WPI; 2002-040725/05.

DR P-PSDB; ABB55784.

XX New secreted proteins and encoding polynucleotides, useful in gene  
therapies, particularly for preventing or treating autoimmune  
disorders, cancer, graft-versus-host disease, wound, osteoporosis,  
stroke or inflammations

PS Disclosure; Page 310-311; 349pp; English.

XX The invention relates to isolated polynucleotides (ABA90876-ABA90968 and  
ABA90980) and encoded proteins (ABB55698-ABB55800), especially  
polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA50885) and  
proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in  
clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1  
are deposited with the American Type Culture Collection (ATCC) with  
accession number 98599. The polynucleotides and encoded polypeptides have  
cytostatic, anti-inflammatory, immunomodulatory, vulnary,  
neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic  
and anti-inflammatory activity and acting as cytokine modulators,  
haematopoiesis regulators, tissue growth modulators and/or cadherin  
suppressors. The polypeptides and polynucleotides are useful in gene  
therapies, particularly for preventing, treating or ameliorating any of  
the following diseases: immune deficiency and disorders; e.g. bacterial  
or fungal infections, autoimmune disorders, cancer, systemic lupus  
erythematosus or graft-versus-host disease; myeloid or lymphoid cell  
deficiencies; wound, burns, incisions and ulcers, osteoporosis or  
osteoarthritis; central and peripheral nervous system diseases and  
neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's  
disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;  
haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis  
or systemic inflammatory response syndrome, ischaemia-reperfusion  
injury, endotoxin lethality, arthritis, inflammatory bowel disease or  
Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus  
foliaceus.

SQ Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 other;

Query Match 19.8%; Score 118.4; DB 24; Length 1534;  
Best Local Similarity 58.9%; Pred. No. 7.7e-25;  
Matches 225; Conservative 0; Mismatches 151; Indels 6; Gaps 1;

QY 158 ACGGTGCCAGTGGAAATGCTCTTCCTCCGCTACATCTCGAGTGGGACCTAGACCTGTG 217  
| ||| |||| | | |||| | | |||| | | |||| | | |||| | | |||| | | |||| |  
Db 811 AAGCGCGCATGCAGAGCTGTGCTTCCTGGAGCTGATTCCTTTTGGGAAGCTGGACCTGG 870  
| |||| | | |||| | | |||| | | |||| | | |||| | | |||| | | |||| | | |||| |  
QY 218 GCGGTGCTACCGGTGCACCTGCTTTCACCTCTCTGGAGGCCCTGCTACGACTGTGCCGCGAC 277







```
FT sig_peptide /product= "Human RMEP-19 protein"
FT 32..94 /*tag= b
FT 95..835
FT mat_peptide /*tag= c
FT /product= "Mature RMEP-19 protein"
XX WO200183524-A2.
XX 08-NOV-2001.
XX 27-APR-2001; 2001WO-US13862.
XX 28-APR-2000; 2000US-200184P.
XX 04-MAY-2000; 2000US-201875P.
XX 06-MAY-2000; 2000US-202090P.
XX 04-JUN-2000; 2000US-210232P.
XX 25-JUL-2000; 2000US-220553P.
XX (INCY-) INCYTE GENOMICS INC.
XX
XX Lal P, Yue H, Tang YT, Lu DAM, Azimzai Y, Au-Young J, Hillman JL;
XX Baughn MR, Yao MG, Burford N, Batra S, Policky JJ;
XX
XX WPI; 2002-034502/04.
XX P-PSDB; AAE15256.
XX
XX New human RNA metabolism protein for diagnosing or treating nervous
XX system disorders, autoimmune/inflammatory disorders, cell proliferative
XX disorders and developmental disorders -
XX
XX Claim 11; Page 180; 196pp; English.
XX
XX The invention relates to human RNA metabolism proteins (RMEP) and their
XX corresponding cDNA molecules. RMEP and its DNA are used for diagnosing,
XX treating and preventing nervous system disorders (epilepsy, dementia,
XX stroke, Alzheimer's disease, Huntington's disease, Parkinson's disease);
XX prion diseases; fatal familial insomnia, nutritional and metabolic
XX diseases of the nervous system; inherited, metabolic, endocrine and
XX toxic myopathy; mental disorders (mood, anxiety, schizophrenic disorders)
XX amnesia and fourrette's disorder; autoimmune/inflammatory disorders (AIDS-
XX acquired immune deficiency syndrome, allergies, anaemia, asthma, gout,
XX atherosclerosis, Crohn's disease, diabetes mellitus, glomerulonephritis,
XX Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid
XX arthritis, osteoporosis, pancreatitis, systemic lupus erythematosus,
XX ulcerative colitis, and infections); cell proliferative disorders (cancer
XX arteriosclerosis, cirrhosis, hepatitis, psoriasis); and developmental
XX disorders (renal tubular acidosis). RMEP DNA is useful in drug screening
XX techniques, gene therapy and for creating transgenic animals. The present
XX sequence is human RMEP-19 cDNA.
XX
XX Sequence 1055 BP; 231 A; 305 C; 270 G; 249 T; 0 other;
XX
XX Query Match 19.5%; Score 116.6; DB 24; Length 1055;
XX Best Local Similarity 58.1%; Pred. No. 2.2e-24;
XX Matches 225; Conservative 0; Mismatches 159; Indels 3; Gaps 1;
XX
XX 153 TAAGACGGCTGCCACGTGGAAATTCCTTCTCGCTACATCTCGGACTGGGACCTAGA 212
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 448 TGAGACCATTTGTCATGCAGAAAGGTGCTTCTCTCTCTGTTCTGTGACACACTGTGC 507
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 213 CCCTGGCCGCTGCTACCGCGTCACTGCTTCACTCCCTGGAGCCCTGCTACGACTGTGC 272
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 508 TCCTAACACAACATACGAGGTACCTGTGACACATCTTGAGCCCTTGCCACAGATGTGC 567
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 273 CCGACATGTGCCGACTTTCTGCGAGGGAACCCCAACCTAGCTGAGGATCTTCACCGC 332
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 568 AGGGAGGTGGCCGAGTTCTTGGCCAGGCAGCAACGTGAATCTCACCATCTTCACCGC 627
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 333 GCGCCTCTACTTCTGTGAGGACCGCAGGCTGAGCCGAGGGCTGCGCGGCTGCACCG 392
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 628 CCGCCTCTGCTACTTCTGTTGGGA---TACAGATTACCAGGAGGGGCTCTGCAGCCTGAGTCA 684
```

```
QY 393 CGCCGGGTGCAAAATAGCCATCATGACCTTCAAGATTTATTTTACTCTCTGGAATACCTTT 452
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
685 GGAAGGGCCCTCCGTGAAGATCATGGGCTACAAAGATTTTGTATCTTGTGAAAAACTT 744
QY 453 TGTAGAAAACCATGAAGAAGCTTTCAAGAGCTTGGAGGGCTGCATGAAATTCAGTTCCG 512
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
745 TGTGTACAGTATGATGAGCCATTCAAGCCCTTGAAGGAGGACTACAAACCAACTTTCGACT 804
QY 513 TCTCTCCAGACAGCTTGGGCGCATCCT 539
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
805 TCTGAAAAGAGGCTACGGGAGATTCT 831

RESULT 12
AAV48231
ID AAV48231 standard; cDNA; 610 BP.
XX
XX AC AAV48231;
XX
XX 09-NOV-1998 (first entry)
XX
XX Human RNA editing enzyme nucleotide sequence.
DE
XX
XX ss; human; RNA editing enzyme; REE; pharmaceutical carrier; cancer;
XX viral disease; circulatory system disorder; RNA processing;
XX hypercholesterolaemia; alpha-galactosidase; apolipoprotein B.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FH 1..573
FT /*tag= a
FT /product= "RNA editing enzyme"
XX
XX US5804185-A.
XX
XX 08-SEP-1998.
XX
XX 13-MAR-1997; 97US-0816241.
XX
XX 13-MAR-1997; 97US-0816241.
XX (INCY-) INCYTE PHARM INC.
XX Bandman O, Goli SK;
XX
XX WPI; 1998-505585/43.
XX P-PSDB; AAW77092.
XX
XX Human RNA editing enzyme and poly.nucleotide(s) encoding it - useful
XX for recombinant production of the enzyme and treatment and detection
XX of disorders associated with incorrect RNA processing
XX
XX Claim 4; Fig 1; 27pp; English.
XX
XX The human RNA editing enzyme (REE) is used in a pharmaceutical carrier
XX for the treatment of cancer, viral diseases and circulatory system
XX disorders. The enzyme is used in vivo for the correct processing of RNA
XX transcripts of genes e.g. change of a codon in apolipoprotein B (apob)
XX RNA to give a 100 and 48 kda product transcribed from the same
XX gene. Certain disorders have been linked to incorrect RNA editing, e.g.
XX failure of apo B editing leads to excessive apob 100 production and
XX hypercholesterolaemia. Other disorders thought to be linked to incorrect
XX RNA processing include aberrant alpha-galactosidase processing in
XX Fabry's disease and neurofibromatosis type I. The enzyme can be produced
XX recombinantly to treat related disorders. It can also be used to raise
XX antibodies for immuno-based detection of REE expression levels
XX e.g. ELISA.
XX
XX Sequence 610 BP; 158 A; 154 C; 154 G; 144 T; 0 other;
XX
XX Query Match 19.2%; Score 114.8; DB 19; Length 610;
XX Best Local Similarity 53.9%; Pred. No. 5.9e-24;
```

Matches 289; Conservative 0; Mismatches 232; Indels 15; Gaps 2;

QY 16 ATGAACCGGAGGAGTTCTTTACCAATTCACAAATGTCGGCTAAGGTCGGCGT 75  
 DB 34 ATGATATCCAGGCACATCTTCTTCCAAATTTAAACACCTATGGAAAGCCACATCGGAAC 93  
 QY 76 GAGACCTACTGTGCTTACCTAGTAGAGAGCGCTGACAGTGTACATCTTTTCACTGGAC 135  
 DB 94 GAAACTTGGCTGTCTTCCCGTGGAGGTATATAAGCGCGCTCAGTCTCTCTCTGGAAG 153  
 QY 136 TTTGGTTATCTTCGCAA-----TAGAACGGCTGGCCACGTGAATTTGCTCTTC 183  
 DB 154 ACGGGCGTCTCGAAACAGGTGGATTCTGAGACCCATTGTATCGAAGAGGTCTTC 213  
 QY 184 CTCGGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGTACCCGCTACCTGGTTC 243  
 DB 214 CTCTTTGGTTCTTGCGACGACATCTGTCTCTTAACACAAAGTACCAAGTCACTGGTAC 273  
 QY 244 ACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGACTTTTTCGAGGGAAC 303  
 DB 274 ACATCTTGGAGCCCTTGGCCAGACTGTGAGGGGAGGTGGCGAGTTCTTGGCCAGGCAC 333  
 QY 304 CCACACTCAGTGTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCCCAAGGCT 363  
 DB 334 AGCAAGTGAATCTCACCATCTTCCACCGCGCGCTCTCTACTTCCAGTATC---CATGT 390  
 QY 364 GAGCCGAGGGCTGCGGGCGCTGCACCGCGCGGTGCAAAATGACCATCATGACCTTC 423  
 DB 391 TACCAGAGGGCTCCGACGCTGAGTCAGGAAGGGTGCCTGTGGAGTATGGAATAT 450  
 QY 424 AAAGATTATTTTACTGCTGGAATATCTTTTGTAGAAACCATGAAAGAACTTTCAAGCC 483  
 DB 451 GAAGATTTTAAATATTGTTGGGAAACTTTGTGTACAATGATAATGAGCCATTCAAGCCT 510  
 QY 484 TGGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCCCATCCT 539  
 DB 511 TGGGAAGGGATTAAACCAACTTTTCGACTTCTGAAAAGAGGCTACGGGAGAGTCT 566

RESULT 13  
 AA72058  
 ID AA72058 standard; cdna; 610 BP.  
 AC AA72058;  
 XX (first entry)  
 DT 24-NOV-2000  
 XX cdna encoding human RNA editing enzyme REE-2.  
 DE RNA editing enzyme; REE-2; human; HEPR homologue; REPR homologue;  
 KW phorbol I homologue; cancer; tumour; autoimmune disorder;  
 KW circulatory system disorder; hypercholesterolaemia; viral infection;  
 KW neurological disease; neurofibromatosis; transcript editing; detection;  
 KW ss.  
 XX Homo sapiens.  
 OS  
 XX Key Location/Qualifiers  
 FH 1..573  
 FT /\*tag= a  
 FT /\*product= "Human REE-2"  
 FT  
 FN US6087108-A.  
 XX  
 XX 11-JUL-2000.  
 PD  
 XX 03-AUG-1998; 98US-0128395.  
 FF  
 XX 13-MAR-1997; 97US-0816241.  
 PR  
 XX (INCY-) INCYTE PHARM INC.  
 PA  
 XX Gol1 SK, Bandman O;  
 PI

XX WPI; 2000-531340/48.  
 DR P-PSDB; AAB11973.  
 XX  
 PT Detecting polynucleotide encoding human RNA editing enzyme comprising  
 PT hybridizing an isolated and purified polynucleotide complementary to  
 PT the polynucleotide and detecting the hybridization complex  
 XX  
 PS Claim 5; Fig 1A-B; 27pp; English.  
 XX This sequence represents the cdna encoding human RNA editing enzyme  
 CC REE-2. cdna encoding REE-2 was initially isolated in a prostate tumour  
 CC cdna library, with the present sequence representing a consensus. REE-2  
 CC has chemical and structural homology with the human apob mRNA editing  
 CC protein HEPR (28% identity), the rat HEPR homologue REPR (30% identity),  
 CC and a portion of the mRNA editing enzyme phorbol I (43% identity).  
 CC REE-2 was found to be expressed in a variety of cdna libraries, a high  
 CC proportion of which were derived from tumours, neuronal tissues, immune  
 CC system cells or synovial tissue from arthritis patients. REE-2 is  
 CC therefore thought to be associated with the development of cancer,  
 CC hypercholesterolaemia), viral infections and neurological diseases (e.g.,  
 CC neurofibromatosis). REE-2 or its nucleic acids may be used in the  
 CC diagnosis, treatment and prevention of such diseases via the modulation  
 CC of transcript editing, which in turn has effects on the encoded protein  
 CC (e.g., an alteration in protein activity). The invention specifically  
 CC relates to methods of detecting nucleic acids encoding human REE-2 in a  
 CC biological sample.  
 XX  
 SQ Sequence 610 BP; 158 A; 154 C; 154 G; 144 T; 0 other;  
 Query Match 19.2%; Score 114.8; DB 21; Length 610;  
 Best Local Similarity 53.9%; Pred. No. 5.9e-24;  
 Matches 289; Conservative 0; Mismatches 232; Indels 15; Gaps 2;

QY 16 ATGAACCGGAGGAGTTCTTTACCAATTCACAAATGTCGGCTAAGGTCGGCGT 75  
 DB 34 ATGATATCCAGGCACATCTTCTTCCAAATTTAAACACCTATGGAAAGCCACATCGGAAC 93  
 QY 76 GAGACCTACTGTGCTTACGCTAGTAGAGAGCGGTGACAGTGTACATCTTTTCACTGGAC 135  
 DB 94 GAAACTTGGCTGTGCTTCCCGTGGAGGTATATAAGCGCGCTCAGTCTCTCTCTGGAAG 153  
 QY 136 TTTGGTTATCTTCGCAA-----TAGAACGGCTGGCCACGTGAATTTGCTCTTC 183  
 DB 154 ACGGGCGTCTCGAAACAGGTGGATTCTGAGACCCATTGTATCGAAGAGGTCTTC 213  
 QY 184 CTCGGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGTACCCGCTACCTGGTTC 243  
 DB 214 CTCTTTGGTTCTTGCGACGACATCTGTCTCTTAACACAAAGTACCAAGTCACTGGTAC 273  
 QY 244 ACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCGACTTTTTCGAGGGAAC 303  
 DB 274 ACATCTTGGAGCCCTTGGCCAGACTGTGAGGGGAGGTGGCGAGTTCTTGGCCAGGCAC 333  
 QY 304 CCACACTCAGTGTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCCCAAGGCT 363  
 DB 334 AGCAAGTGAATCTCACCATCTTCCACCGCGCGCTCTCTACTTCCAGTATC---CATGT 390  
 QY 364 GAGCCGAGGGCTGCGGGCGCTGCACCGCGCGGTGCAAAATGACCATCATGACCTTC 423  
 DB 391 TACCAGAGGGCTCCGACGCTGAGTCAGGAAGGGTGCCTGTGGAGTATGGAATAT 450  
 QY 424 AAAGATTATTTTACTGCTGGAATATCTTTTGTAGAAACCATGAAAGAACTTTCAAGCC 483  
 DB 451 GAAGATTTTAAATATTGTTGGGAAACTTTGTGTACAATGATAATGAGCCATTCAAGCCT 510  
 QY 484 TGGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCCCATCCT 539  
 DB 511 TGGGAAGGGATTAAACCAACTTTTCGACTTCTGAAAAGAGGCTACGGGAGAGTCT 566

RESULT 14

AAS41420  
ID AAS41420 standard; cDNA; 950 BP.  
AC AAS41420;  
XX  
DT 17-DEC-2001 (first entry)  
XX  
DE cDNA encoding novel human enzyme polypeptide #636.  
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; gene therapy; cytostatic;  
KW anti arthritic; nephrotropic; anticoagulant; ss.  
OS  
XX Homo sapiens.  
XX WO200155301-A2.  
PN  
XX  
XX  
PD  
XX  
XX  
PF 17-JAN-2001; 2001WO-US01239.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.

PR	17-NOV-2000;	2000US-0249300.	
PR	01-DEC-2000;	2000US-0250160.	
PR	01-DEC-2000;	2000US-0250391.	
PR	05-DEC-2000;	2000US-0251030.	
PR	05-DEC-2000;	2000US-0251988.	
PR	05-DEC-2000;	2000US-0256719.	
PR	06-DEC-2000;	2000US-0251479.	
PR	08-DEC-2000;	2000US-0251856.	
PR	08-DEC-2000;	2000US-0251868.	
PR	08-DEC-2000;	2000US-0251869.	
PR	08-DEC-2000;	2000US-0251989.	
PR	11-DEC-2000;	2000US-0251990.	
PR	11-DEC-2000;	2000US-0254097.	
PR	05-JAN-2001;	2001US-0259678.	
XX	(HUMA-) HUMAN GENOME SCI INC.		
XX	Rosen CA, Barash SC, Ruben SM;		
XX	WPI; 2001-465566/50.		
DR	P-PSDB; AAU23550.		
XX	Novel polypeptides and polynucleotides useful for diagnosing,		
PT	preventing, treating neural, immune system, muscular, reproductive,		
PT	pulmonary, cardiovascular, renal, proliferative disorders and cancerous		
PT	diseases -		
XX	Claim 4; SEQ ID No 646; 1180pp; English.		
XX	The present invention relates to the isolation of novel human enzyme		
XX	polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences		
CC	encoding them. The enzyme polypeptides of the invention may comprise the		
CC	functional classes of oxidoreductases, transferases, hydrolases, lyases,		
CC	isomerases or ligases. The sequences of the invention are useful in the		
CC	diagnosis, treatment, prevention and/or prognosis of a wide range of		
CC	disorders including hyperproliferative disorders (e.g. cancer),		
CC	immunodeficiency disorders (e.g. AIDS) autoimmune disorders		
CC	(e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),		
CC	metabolic disorders (e.g. phenylketonuria), inflammatory disorders		
CC	(e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),		
CC	blood-related disorders (e.g. haemophilia), reproductive disorders		
CC	(e.g. infertility) and infectious disorders (e.g. influenza). The		
CC	polynucleotides of the invention can also be used in gene therapy.		
CC	AAU40785-AAU41684 represent cDNA sequences encoding for the novel human		
CC	enzyme polypeptides of the invention.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp.wipo.int/pub/published_pct_sequences.		
XX	Sequence 950 BP; 219 A; 289 C; 219 G; 220 T; 3 other;		
SQ	Query Match 19.2%; Score 114.8; DB 22; Length 950;		
	Best Local Similarity 53.9%; Pred. No. 7.3e-24;		
	Matches 289; Conservative 0; Mismatches 232; Indels 15; Gaps 2;		
QY	16 ATGAACCGGAGGAGTTCTTTTACCAATTCAAAATGTCGCGGTAAAGGTGCGCGT 75		
DB	95 ATGTATCCAGGCACATTCTACTTCCATTTTAAAAACCTATGGAAGCCACGATCGGAAC 154		
QY	76 GAGACCTACCTGTGCTAGTAGTGAGAGCGGTGACAGTGTACATCTTTTCACTGGAC 135		
DB	155 GAAACTTGGCTGTCTTCCACCGTGGAGGATATAAGCGCGCGCTCAGTTGCTCTGGAAG 214		
QY	136 TTTTGGTTATCTTCCAA-----TAAGAACGGCTGCCACGTGGAAATGCTCTTC 183		
DB	215 ACGGGCGTCTCCGAACAGGTGGATTCTGAGACCCATTGTCTATGAGAAAGGTGCTTC 274		
QY	184 CTCGCGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTCTACCGGTACCTGGTTC 243		
DB	275 CTCCTTGTGTCGACGACATATCTCTCTTAACACAAAGTACAGGTACCTGGTAC 334		
QY	244 ACCCTCTGAGCCCTGCTACGACTGTGCCGACATGTGGCGCGACTTTCTCGGAGGAAC 303		

Db	335 ACATCTTGGAGCCCTTCCACAGACTGTGCAGGGAGGTGCCGAGTTCTTGGCCAGGCAC 394
Qy	304 CCAACCTCAGTCTGAGGATCTTCCAGCGCGCTCTACTTCTGTGAGGACCGCAAGGCT 363
Db	395 AGCAAGCTGAATCTCACCATCTTTCACGCGCGCTCTTACTACTTCCAGTATC---CATGT 451
Qy	364 GAGCCCGAGGGCTGGCGGCTGCACCGCGCGGTGCAAAATAGCCATCATGACCTTC 423
Db	452 TACCAGGAGGGCTCCCGACGCTGAGTCAGAAAGGGTCTGCTGGAGATCATGCACATAT 511
Qy	424 AAAGATTATTTTACTCTGTAATACATTTTGTAGAAACCATGAAACATTTTCAAGGCC 483
Db	512 GAAGATTAAATATTGTTGGGAAACCTTTTGTGTACAAATGATATGAGCCATTCAAGCCT 571
Qy	484 TGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCT 539
Db	572 TGAAGGGATTAAAAACCACTTTTCGACTTCTGAAAAGAGGCTACGGGAGAGTCT 627
RESULT 15	
AAF16264	
ID	AAF16264 standard; cDNA: 987 BP.
XX	AC AAF16264;
XX	13-MAR-2001 (first entry)
DT	Human prostate cancer antigen nucleotide sequence SEQ ID NO:699.
DE	
XX	Human prostate cancer; prostate cancer antigen; detection; diagnosis;
KW	neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;
KW	vulnary; gastrointestinal; nephrotropic; antineoplastic; gynaecological;
KW	antibacterial; gene therapy; neural; immune; reproductive; renal;
KW	gastrointestinal; pulmonary; cardiovascular; proliferative disorder;
KW	wound; infectious disease; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO200055174-A1.
XX	
PD	21-SEP-2000.
XX	
PF	08-MAR-2000; 2000WO-US05988.
XX	
PR	12-MAR-1999; 99US-0124270.
XX	
XX	(HUMA-) HUMAN GENOME SCI INC.
PA	(ROSE/) ROSEN C A.
XX	
PI	Rosen CA, Ruben SM;
XX	
DR	WPI; 2000-587513/55.
DR	P-PSDB; AAB57061.
XX	
PT	Prostate cancer associated gene sequences, referred to as prostate
PT	cancer antigens, useful for treatment, prevention, and diagnosis of
PT	disorders such as prostate cancer -
XX	
PS	Claim 1; Page 1129; 2338pp; English.
XX	
CC	AAF15566 to AAF16505 encode the human prostate cancer associated
CC	proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.
CC	The prostate cancer antigens can have neuroprotective, cytostatic,
CC	cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,
CC	nephrotropic, antineoplastic, gynaecological and antibacterial activities,
CC	and can be used in gene therapy. The prostate cancer antigen
CC	polynucleotides may be used for detection of prostate cancer, chromosome
CC	identification, as chromosome markers, and for numerous other diagnostic
CC	or research purposes. The prostate cancer antigens may be used to treat
CC	disorders such as neural, immune, muscular, reproductive,
CC	gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC	disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to
CC	AAB57303 represent sequences used in the exemplification of the present





GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 23:34:44 ; Search time 805.004 Seconds  
(without alignments)  
12010.756 Million cell updates/sec

Title: US-09-966-880A-7 COPY 80 676

perfect score:

Sequence: 1 atggacagcctcttgatgaa.....ttcgactttgggactttga 597

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 sears. 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : EST:★

EST: \*

```
1: em_estba:*
```

2: em\_esthum:

3: em\_estin:\*

4: em\_estmu: \*

5: em\_estov:\*

6: em\_estpl:\*

```
7: em_estro:*
```

```
8: em_hlc:*
```

9: gb\_est1:\*

```
10: gb_est2:*
```

```
11: gb_htc:*
```

```
12: gb_est3:*
```

```
13: gb_est4:*
14: gb_est4:*
```

14: gb\_est5: \*

15: em\_estun:

```
16: em_estom: *
17: sb_acc: *
```

```
17: go_gss: *
18: em acc bun
```

18: em\_gss\_1111  
19: em\_gss\_111119: em\_gss\_inval  
20: em\_gss\_not

```
20: em_gss_ptr
21: em_gss_vrt
```

21: em\_gss\_v10

23: em\_qss\_max

24: em\_gss\_mus

25: em\_gss\_oth

26: em\_gss\_pro

27: em\_gss\_root

# 6

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	597	100.0	743	12	BG686133	BG686133 602638412
2	597	100.0	856	9	AL559877	AL559877 AL559877
3	597	100.0	872	12	BG758510	BG758510 602712721
4	597	100.0	953	14	BQ655440	BQ655440 AGENCOURT
5	597	100.0	1052	14	BQ055935	BQ055935 AGENCOURT
6	581.8	97.5	820	12	BG757089	BG757089 602715124

## ALIGNMENTS

RESULT 1	
LOCUS	BG686133
DEFINITION	602638412F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4766234 5', linear mRNA 743 bp EST 01-MAY-2001
ACCESSION	BG686133
VERSION	BG686133
KEYWORDS	EST.
SOURCE	human..
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 743)
TITLE	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-femail.nih.gov">cgapbs-femail.nih.gov</a> Tissue Procurement: Louis M. Staudt, M.D., Ph.D. cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLCMI626 row: 9 column: 03 High quality sequence stop: 740. Location/Qualifiers 1..743
FEATURES	
SOURCE	

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4766234"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pORF7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
BASE COUNT 176 a 197 c 188 g 182 t
ORIGIN

```

```

Query Match 100.0%; Score 597; DB 12; Length 743;
Best Local Similarity 100.0%; Pred. No. 1.3e-168;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACAGCCTCTTGATGACCGGAGGAGCTTTCTTACCAATTCAAAATGTCGCTGG 60
DB 78 ATGACAGCCTCTTGATGACCGGAGGAGCTTTCTTACCAATTCAAAATGTCGCTGG 137
QY 61 GCTAAGGGTGGCGGTGAGACCTACCTGTGCTAGTAGTAGAGGCGGTGACAGTGTCTA 120
DB 138 GCTAAGGGTGGCGGTGAGACCTACCTGTGCTAGTAGTAGAGGCGGTGACAGTGTCTA 197
QY 121 TCCTTTTCTAGTGTGTTTATCTCGCAATTAAGAACGGCTGCCACGTGGAATGCTC 180
DB 198 TCCTTTTCTAGTGTGTTTATCTCGCAATTAAGAACGGCTGCCACGTGGAATGCTC 257
QY 181 TTCTTCGCTACATCTCGGACTGGGACTAGACCTGGCGGCTGTACCGCGTCACTGG 240
DB 258 TTCTTCGCTACATCTCGGACTGGGACTAGACCTGGCGGCTGTACCGCGTCACTGG 317
QY 241 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCGAGGG 300
DB 318 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCGAGGG 377
QY 301 AACCCCAACCTCAGTGTAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
DB 378 AACCCCAACCTCAGTGTAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 437
QY 361 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGCTGCATAATAGCCATCATGACC 420
DB 438 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGCTGCATAATAGCCATCATGACC 497
QY 421 TTCAAGATTTATTTTACTGCTGGAATACATTTTGTAGAAACCATGAAAGAACTTTCAA 480
DB 498 TTCAAGATTTATTTTACTGCTGGAATACATTTTGTAGAAACCATGAAAGAACTTTCAA 557
QY 481 GCCTGGGAGGGCTGATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGGCATCCTT 540
DB 558 GCCTGGGAGGGCTGATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGGCATCCTT 617
QY 541 TTGCCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTCTGTTGGGACTTTGA 597
DB 618 TTGCCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTCTGTTGGGACTTTGA 674

```

```

RESULT 2
AL559877 LOCUS AL559877.LTI_FL011_B01 Homo sapiens cDNA clone CS0DG003YB14 5 prime
DEFINITION / mRNA sequence.
ACCESSION AL559877
VERSION AL559877.1 GI:12905793
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

```

```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 856)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..856
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DG003YB14"
/clone_lib="LTI_FL011_B01"
/sex="male"
/tissue_type="B cells from Burkitt lymphoma"
/lab_host="DH10B"
/notes="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned, into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center-Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"
BASE COUNT 209 a 217 c 202 g 226 t 2 others
ORIGIN

```

```

Query Match 100.0%; Score 597; DB 9; Length 856;
Best Local Similarity 100.0%; Pred. No. 1.4e-168;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACAGCCTCTTGATGACCGGAGGAGCTTTCTTACCAATTCAAAATGTCGCTGG 60
DB 19 ATGACAGCCTCTTGATGACCGGAGGAGCTTTCTTACCAATTCAAAATGTCGCTGG 78
QY 61 GCTAAGGGTGGCGGTGAGACCTACCTGTGCTAGTAGTAGAGGCGGTGACAGTGTCTA 120
DB 79 GCTAAGGGTGGCGGTGAGACCTACCTGTGCTAGTAGTAGAGGCGGTGACAGTGTCTA 138
QY 121 TCCTTTTCTAGTGTGTTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAATGCTC 180
DB 139 TCCTTTTCTAGTGTGTTTATCTTCGCAATTAAGAACGGCTGCCACGTGGAATGCTC 198
QY 181 TTCTTCGCTACATCTCGGACTGGGACTAGACCTGGCGGCTGTACCGCGTCACTGG 240
DB 199 TTCTTCGCTACATCTCGGACTGGGACTAGACCTGGCGGCTGTACCGCGTCACTGG 258
QY 241 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGCGAGGG 300
DB 259 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGAGCTTTCTGCGAGGG 318
QY 301 AACCCCAACCTCAGTGTGAGGATCTTCACCGCGGCTCTACTTCTGTGAGGACCGCAAG 360
DB 319 AACCCCAACCTCAGTGTGAGGATCTTCACCGCGGCTCTACTTCTGTGAGGACCGCAAG 378
QY 361 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGCTGCATAATAGCCATCATGACC 420
DB 379 GCTGAGCCCGAGGGGCTGCGCGGCTGCACCGCGCGGCTGCATAATAGCCATCATGACC 438
QY 421 TTCAAGATTTATTTTACTGCTGGAATACATTTTGTAGAAACCATGAAAGAACTTTCAA 480
DB 439 TTCAAGATTTATTTTACTGCTGGAATACATTTTGTAGAAACCATGAAAGAACTTTCAA 498
QY 481 GCCTGGGAGGGCTGATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGGCATCCTT 540
DB 499 GCCTGGGAGGGCTGATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGCGGCATCCTT 558
QY 541 TTGCCCTGTATGAGGTTGATGACTTACGAGAGCGCATTTCTGTTGGGACTTTGA 597

```



```
Db 559 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTTCGTACTTTGGGACTTTGA 615

RESULT 3
Bg758510
LOCUS 872 bp mRNA linear EST 15-MAY-2001
DEFINITION 602712721F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4853069 5',
mRNA sequence.
ACCESSION Bg758510
VERSION Bg758510.1 GI:14069163
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 872)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1698 row: i column: 06
High quality sequence stop: 836.
Location/Qualifiers
1. 872
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4853069"
/clone_lib="NIH_MGC_48"
/lab_host="DH10B (phage-resistant)"
/note="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 211 a 221 c 212 g 228 t
ORIGIN
Query Match 100.0%; Score 597; DB 12; Length 872;
Best Local Similarity 100.0%; Pred. No. 1.4e-168;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGGACAGCCTTGTATGACCGGAGGAGTTTCTTTACCAATTCAAAATGTCGCGTGG 60
Db 66 ATGGACAGCCTTGTATGACCGGAGGAGTTTCTTTACCAATTCAAAATGTCGCGTGG 125
QY 61 GCTAGGTCGCGGTGACACCTACCTGTGCTACGTAGTGAAGCGGTGACAGTCTACA 120
Db 126 GCTAAGGTCGCGGTGACACCTACCTGTGCTACGTAGTGAAGCGGTGACAGTCTACA 185
QY 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCAAGTGTGCTC 180
Db 186 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCAAGTGTGCTC 245
QY 181 TTCTCCGCTACATCTCGAGCTGGGACCTAGACCTCGCGCGTGTCTACCGGTTCACCTGG 240
Db 246 TTCTCCGCTACATCTCGAGCTGGGACCTAGACCTCGCGCGTGTCTACCGGTTCACCTGG 305
QY 241 TTCACCTCTGGAGCCCTCTGTACAGCTGTGCCCGACATGTGGCCGACTTTCCTGGAGGG 300
Db 306 TTCACCTCTGGAGCCCTCTGTACAGCTGTGCCCGACATGTGGCCGACTTTCCTGGAGGG 365

QY 301 AACCCCAACCTCAGTCTGAGGATCTTACCCGCGGCGCTCTACTTCTGTGAGGACCGCAAG 360
Db 366 AACCCCAACCTCAGTCTGAGGATCTTACCCGCGGCGCTCTACTTCTGTGAGGACCGCAAG 425
QY 361 GCTGAGCCCGAGGGGCTGCGGGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACC 420
Db 426 GCTGAGCCCGAGGGGCTGCGGGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACC 485
QY 421 TTCAAGATTTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTTCAA 480
Db 486 TTCAAGATTTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTTCAA 545
QY 481 GCCTGGGAAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 546 GCCTGGGAAGGCTGCATGAAATTCAGTCTCTCCAGACAGCTTCGGCGCATCCTT 605
QY 541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTTCGTACTTTGGGACTTTGA 597
Db 606 TTGCCCTGTATGAGTTGATGACTTACGAGACGCAATTTCGTACTTTGGGACTTTGA 662

RESULT 4
BQ065440
LOCUS 953 bp mRNA linear EST 02-APR-2002
DEFINITION AGENCOURT_6855061 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5929977
5', mRNA sequence.
ACCESSION BQ065440
VERSION BQ065440.1 GI:19894486
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 953)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Lou Staudt
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2108 row: p column: 10
High quality sequence stop: 634.
Location/Qualifiers
1. 953
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5929977"
/clone_lib="NIH_MGC_99"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lymph; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 238 a 236 c 233 g 246 t
ORIGIN
Query Match 100.0%; Score 597; DB 14; Length 953;
Best Local Similarity 100.0%; Pred. No. 1.5e-168;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 ATGCAGACGCTCTGTATGTAACCGGAGGAAGTTTCTTTACCAATTCACCAATATGTCCTGG 60
Db 64 ATGCAGACGCTCTGTATGTAACCGGAGGAAGTTTCTTTACCAATTCACCAATATGTCCTGG 123
QY 61 GCTAAGGTCGGCGTCAGACCTACCTGTGCTAGCTAGTGAAGAGGCGTGACAGTGTCTACA 120
Db 124 GCTAAGGTCGGCGTCAGACCTACCTGTGCTAGCTAGTGAAGAGGCGTGACAGTGTCTACA 183
QY 121 TCCTTTTCACTGTGACTTTGGTTATCTTCGCAATTAAGAAGGCTGCGCACGTTGGAATTTGCTC 180
Db 184 TCCTTTTCACTGTGACTTTGGTTATCTTCGCAATTAAGAAGGCTGCGCACGTTGGAATTTGCTC 243
QY 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGTACCGCGTCACCTGG 240
Db 244 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGTACCGCGTCACCTGG 303
QY 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGGCAGGG 300
Db 304 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGGCAGGG 363
QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Db 364 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 423
QY 361 GCTGAGCCGAGGGGTCGGCGGCTGACCGCGCGGCTGCAATAGCCATCATGAC 420
Db 424 GCTGAGCCGAGGGGTCGGCGGCTGACCGCGCGGCTGCAATAGCCATCATGAC 483
QY 421 TTCAAAGATTTATTTTACTGCTGGAATCTTTGTAGAAACCATGAAAGAACTTTCAAA 480
Db 484 TTCAAAGATTTATTTTACTGCTGGAATCTTTGTAGAAACCATGAAAGAACTTTCAAA 543
QY 481 GCCTGGAAAGGGTCGATGAAATTCAGTTTCAGTTCCTCCAGACAGCTTCGGCGCATCCTT 540
Db 544 GCCTGGAAAGGGTCGATGAAATTCAGTTTCAGTTCCTCCAGACAGCTTCGGCGCATCCTT 603
QY 541 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCATTTTCGTAATTTGGGACTTTGA 597
Db 604 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCATTTTCGTAATTTGGGACTTTGA 660
```

```
RESULT 5
BQ055935
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1052)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Lou Staudt
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM2051 row: m column: 14
High quality sequence stop: 665.
Location/Qualifiers
1. .1052
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5808181"
```

FEATURES
source

```
/clone_lib="NIH_MGC_99"
/tissue_type="lymphoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: lymph; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 276 a 250 c 258 g 266 t 2 others
ORIGIN
Query Match 100.0%; Score 597; DB 14; Length 1052;
Best Local Similarity 100.0%; Pred. No. 1.6e-168;
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCAGACGCTCTGTATGTAACCGGAGGAAGTTTCTTTACCAATTCACCAATATGTCCTGG 60
Db 64 ATGCAGACGCTCTGTATGTAACCGGAGGAAGTTTCTTTACCAATTCACCAATATGTCCTGG 123
QY 61 GCTAAGGTCGGCGTCAGACCTACCTGTGCTAGCTAGTGAAGAGGCGTGACAGTGTCTACA 120
Db 124 GCTAAGGTCGGCGTCAGACCTACCTGTGCTAGCTAGTGAAGAGGCGTGACAGTGTCTACA 183
QY 121 TCCTTTTCACTGTGACTTTGGTTATCTTCGCAATTAAGAAGGCTGCGCACGTTGGAATTTGCTC 180
Db 184 TCCTTTTCACTGTGACTTTGGTTATCTTCGCAATTAAGAAGGCTGCGCACGTTGGAATTTGCTC 243
QY 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGTACCGCGTCACCTGG 240
Db 244 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGTACCGCGTCACCTGG 303
QY 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGGCAGGG 300
Db 304 TTCACCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGGCAGGG 363
QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
Db 364 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 423
QY 361 GCTGAGCCGAGGGGTCGGCGGCTGACCGCGCGGCTGCAATAGCCATCATGAC 420
Db 424 GCTGAGCCGAGGGGTCGGCGGCTGACCGCGCGGCTGCAATAGCCATCATGAC 483
QY 421 TTCAAAGATTTATTTTACTGCTGGAATCTTTGTAGAAACCATGAAAGAACTTTCAAA 480
Db 484 TTCAAAGATTTATTTTACTGCTGGAATCTTTGTAGAAACCATGAAAGAACTTTCAAA 543
QY 481 GCCTGGAAAGGGTCGATGAAATTCAGTTTCAGTTCCTCCAGACAGCTTCGGCGCATCCTT 540
Db 544 GCCTGGAAAGGGTCGATGAAATTCAGTTTCAGTTCCTCCAGACAGCTTCGGCGCATCCTT 603
QY 541 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCATTTTCGTAATTTGGGACTTTGA 597
Db 604 TTGCCCTCTATGAGGTTGATGACTTACGAGACGCATTTTCGTAATTTGGGACTTTGA 660
RESULT 6
BQ757089
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 820)
602715124F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4855517 5',
mrna sequence.
820 bp mrna linear EST 15-MAY-2001
BQ757089
BQ757089.1 GI:14067742
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 820)
```

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-re@mail.nih.gov](mailto:cgabbs-re@mail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCMI704 row: o column: 06  
High quality sequence stop: 675.  
Location/Qualifiers  
1. 820  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4855517"  
/clone\_lib="NIH\_MGC\_48"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dr priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."  
210 a 230 c 200 g 180 t

#### FEATURES

source

BASE COUNT 210 a 230 c 200 g 180 t  
ORIGIN  
Query Match 97.5%; Score 581.8; DB 12; Length 820;  
Best Local Similarity 99.5%; Pred. No. 5.2e-164;  
Matches 594; Conservative 0; Mismatches 2; Indels 1; Gaps 1;  
QY 1 ATGGACACGCTCTTGATGAACGGAGGAAGTTCTTTACCAATTCAAAATGTCGCGTGG 60  
DB 85 ATGGACACGCTCTTGATGAACGGAGGAAGTTCTTTACCAATTCAAAATGTCGCGTGG 144  
QY 61 GCTAAGGTCGGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 120  
DB 145 GCTAAGGTCGGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 204  
QY 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGGAAATGCTC 180  
DB 205 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGGAAATGCTC 264  
QY 181 TTCTCCGCTACATCTCGGACTGGGACCTAGACCCTGACCCGTGGCGGTGCTACCGTACCTGG 240  
DB 265 TTCTCCGCTACATCTCGGACTGGGACCTAGACCCTGACCCGTGGCGGTGCTACCGTACCTGG 324  
QY 241 TTCACTCTCTGAGCCCTGCTACGACTGTGCCCACATGTGGCCGACTTTCTGCGAGGG 300  
DB 325 TTCACTCTCTGAGCCCTGCTACGACTGTGCCCACATGTGGCCGACTTTCTGCGAGGG 384  
QY 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGGCCCTCTACTTCTGTGAGAGCCGCAAG 360  
DB 385 AACCCCAACCTCAGTCTGAGGATCTTACCGCGGCCCTCTACTTCTGTGAGAGCCGCAAG 444  
QY 361 GCTGAGCCGAGGGGCTGGCGGCTGCACCGCGCGGGGTGCAATAGCCATCATGACC 420  
DB 445 GCTGAGCCGAGGGGCTGGCGGCTGCACCGCGCGGGGTGCAATAGCCATCATGACC 504  
QY 421 TTCAAGATATTTTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATCTTCAAA 480  
DB 505 TTCAAGATATTTTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATCTTCAAA 564  
QY 481 GCTGGGAGGGGCTGCATGAATAATTCAGTTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 540  
DB 565 GCTGGGAGGGGCTGCATGAATAATTCAGTTTCGTCTCTCCAGACAGCTTCGGCGCATCCTT 624

QY 541 TTGCCCTGTATGAGTTGATGACTTAGGACGATTTAGGACGATTTGCTACTTTGGACTTTGA 597  
DB 625 TTGCCCTGTATGAGTTGATGACTTAGGACGATTTAGGACGATTTGCTACTTAGGACTTGA 680  
RESULT 7  
BG757392  
LOCUS 602711022F1 NIH\_MGC\_48 Homo sapiens cDNA clone linear EST 15-MAY-2001  
DEFINITION mRNA sequence.  
ACCESSION BG757392  
VERSION BG757392.1 GI:14068045  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 693)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-re@mail.nih.gov](mailto:cgabbs-re@mail.nih.gov)  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: LLCMI694 row: k column: 05  
High quality sequence stop: 693.  
Location/Qualifiers  
1. 693  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4851580"  
/clone\_lib="NIH\_MGC\_48"  
/tissue\_type="primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dr priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."  
157 a 188 c 178 g 170 t

#### FEATURES

source

BASE COUNT 157 a 188 c 178 g 170 t  
ORIGIN  
Query Match 96.5%; Score 576.4; DB 12; Length 693;  
Best Local Similarity 99.7%; Pred. No. 2e-162;  
Matches 588; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
QY 1 ATGGACACGCTCTTGATGAACGGAGGAAGTTCTTTACCAATTCAAAATGTCGCGTGG 60  
DB 81 ATGGACACGCTCTTGATGAACGGAGGAAGTTCTTTACCAATTCAAAATGTCGCGTGG 140  
QY 61 GCTAAGGTCGGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTAC 119  
DB 141 GCTAAGGTCGGCGGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTAC 200  
QY 120 ATCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGGAAATGCT 179  
DB 201 ATCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAGAAGCGGTGCCAGTGGAAATGCT 260  
QY 180 CTTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGCGCCCTGCTACCGCGTCACTTG 239  
DB 261 CTTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGCGCCCTGCTACCGCGTCACTTG 320

QY 240 GTTCACTCTCTGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGCGAGG 299  
 |||||  
 Db 321 GTTCACTCTCTGAGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTTCTGCGAGG 380  
 QY 300 GAACCCCAACCTAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAA 359  
 Db 381 GAACCCCAACCTAGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAA 440  
 QY 360 GGCTGAGCCCGAGGGCTGCGCGCTGCACCGCGCGGGTGGAATAGCCATCATGAC 419  
 Db 441 GGCTGAGCCCGAGGGCTGCGCGCTGCACCGCGCGGGTGGAATAGCCATCATGAC 500  
 QY 420 CTTCAAGATATTTTACTGCTGAATACTTTGTAGAAAACCATGAAGAACTTTCAA 479  
 Db 501 CTTCAAGATATTTTACTGCTGAATACTTTGTAGAAAACCATGAAGAACTTTCAA 560  
 QY 480 AGCTGGGAAGGGCTGCATGAATATTCAGTTCCTCCAGACAGCTTCCGCGCATCCT 539  
 Db 561 AGCTGGGAAGGGCTGCATGAATATTCAGTTCCTCCAGACAGCTTCCGCGCATCCT 620  
 QY 540 TTTGCCCTGTATGAGGTTGATGACTTACGAGACGCATTTCTGCTATTGG 589  
 |||||  
 Db 621 TTTGCCCTGTATGAGGTTGATGACTTACGAGACGCATTTCTGCTATTGG 670

## RESULT 8

BF975166

LOCUS

DEFINITION 602244657F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4335639 5',  
 mRNA sequence.

ACCESSION

BF975166

VERSION

EST.

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 942)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LNCMI207 row: a column: 16

High quality sequence stop: 707.

Location/Qualifiers

1..942

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:4335639"

/clone\_lib="NIH\_MGC\_48"

/tissue\_type="primary B-cells from tonsils (cell line)"

/lab\_host="DH10B (phage-resistant)"

/notes="Organ: B-cells; Vector: pOTB7; Site:1: xhoI;

Site:2: EcoRI; cDNA made by oligo-dT priming.

Directionally cloned into EcoRI/XhoI sites using the

following 5' adaptor: GGCACGAG(G). Size-selected &gt;500bp

for average insert size 1.8kb. Library constructed by Ling

Hong in the laboratory of Gerald M. Rubin (University of

California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).

Note: this is a NIH\_MGC Library."

249 a 209 c 251 g 233 t

BASE COUNT

ORIGIN

Query Match

93.9%; Score 560.6; DB 12; Length 942;

Best Local Similarity 99.0%; Pred. No. 1.4e-157;

Matches 585; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 1 ATGGACAGCCCTTTCATGAACCGGAGGAGTTTCTTTTACCAA-TTCAAAAATGTCCGCTG 59

Db 78 ATGGACAGCCCTTTCATGAACCGGAGGAGTTTCTTTTACCAAATGTCCGCTG 137

QY 60 GCCTAAGGTCGGCGTGAGACTACCTGTCTACGTAGTGAAGAGCCTGACAGTGTCTAC 119

Db 138 GCCTAAGGTCGGCGTGAGACTACCTGTCTACGTAGTGAAGAGCCTGACAGTGTCTAC 197

QY 120 ATCTTTTCACTGGACTTTGTTATCTTCGCAATAAGAACGCTGCCACGTGGGAATTCCT 179

Db 198 ATCTTTTCACTGGACTTTGTTATCTTCGCAATAAGAACGCTGCCACGTGGGAATTCCT 257

QY 180 CTTCTCCGCTACATCTCGGACTGGGACTAGACCCCTGGCGCTGTACCCGCTACCTGT 239

Db 258 CTTCTCCGCTACATCTCGGACTGGGACTAGACCCCTGGCGCTGTACCCGCTACCTGT 317

QY 240 GTTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGCTGTCTTCTGAGGACCCGCAA 359

Db 318 GTTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGCTGTCTTCTGAGGACCCGCAA 437

QY 360 GGCTGAGCCCGAGGGCTGCGCGCTGCACCGCGCGGGTGCAATATAGCATCATGAC 419

Db 438 GGCTGAGCCCGAGGGCTGCGCGCTGCACCGCGCGGGTGCAATATAGCATCATGAC 497

QY 420 CTTCAAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTCAA 479

Db 498 CTTCAAAGATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTCAA 557

QY 480 AGCTGGGAAGGGCTGCATGAATATTCAGTTCCTCTCCAGACAGCTTCCGCGCATCCT 539

Db 558 AGCTGGGAAGGGCTGCATGAATATTCAGTTCCTCTCCAGACAGCTTCCGCGCATCCT 617

QY 540 TTTGCCCTGTATGA-GGTTGATGACTTACGAGACGCATTTCTGCTACTTTGG 589

Db 618 TTGGGCCCTGTATGAGGTTGATGACTTACGAGACGCATTTCTGCTACTTTGG 668

RESULT 9

BF238155

LOCUS

DEFINITION 601811880F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4054915 5',  
 mRNA sequence.

ACCESSION

BF238155

VERSION

EST.

KEYWORDS

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 541)

NIH-MGC http://mgi.nci.nih.gov/.

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-re@mail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LNCM895 row: p column: 20

High quality sequence stop: 541.

Location/Qualifiers

1..541

FEATURES

source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4054915"  
/clone\_lib="NIH\_MGC\_48"  
/tissue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."  
120 a 150 c 142 g 129 t

BASE COUNT 120 a 150 c 142 g 129 t  
ORIGIN  
Query Match 77.7%; Score 464; DB 12; Length 541;  
Best Local Similarity 100.0%; Pred. No. 1.3e-128;  
Matches 464; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGCCCTGG 60  
DB 78 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGCCCTGG 137  
QY 61 GCTAAGGGTGGCGTGAGACCTACCTGCTAGTAGTGAAGAGGGGTGACAGTGTCTACA 120  
DB 138 GCTAAGGGTGGCGTGAGACCTACCTGCTAGTAGTGAAGAGGGGTGACAGTGTCTACA 197  
QY 121 TCCTTTTACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACCTGGAATTCGTC 180  
DB 198 TCCTTTTACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACCTGGAATTCGTC 257  
QY 181 TTCCTCCGTACATCTCGACCTGGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACTGG 240  
DB 258 TTCCTCCGTACATCTCGACCTGGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACTGG 317  
QY 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACCTTTCTGCGAGGG 300  
DB 318 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACCTTTCTGCGAGGG 377  
QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
DB 378 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 437  
QY 361 GCTGAGCCCGAGGGCTGGCGGGCTGCACCGCGCGGGTGCAGATAGCCATCATGACC 420  
DB 438 GCTGAGCCCGAGGGCTGGCGGGCTGCACCGCGCGGGTGCAGATAGCCATCATGACC 497  
QY 421 TTCAAGATATTTTACTGCTGGAAATCTTTGTAGAAAACCA 464  
DB 498 TTCAAGATATTTTACTGCTGGAAATCTTTGTAGAAAACCA 541

RESULT 10  
BG686876 889 bp mRNA linear EST 01-MAY-2001  
LOCUS 602650861F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4763247 5',  
DEFINITION mRNA sequence.  
ACCESSION BG686876  
VERSION BG686876.1 GI:13918273  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 889)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
CONTACT: Robert Strausberg, Ph.D.  
COMMENT Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1618 row: j column: 16  
High quality sequence start: 6  
High quality sequence stop: 727.  
FEATURES  
Location/Qualifiers  
1..889  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4763247"  
/clone\_lib="NIH\_MGC\_48"  
/tissue\_type="Primary B-cells from tonsils (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. Library constructed by Ling  
Hong in the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies).  
Note: this is a NIH\_MGC Library."  
205 a 223 c 257 g 203 t 1 others  
BASE COUNT 205 a 223 c 257 g 203 t 1 others  
ORIGIN

Query Match 71.6%; Score 427.4; DB 12; Length 889;  
Best Local Similarity 99.8%; Pred. No. 1.8e-117;  
Matches 428; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGCCCTGG 60  
DB 38 ATGGACAGCCTCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAATGCCCTGG 97  
QY 61 GCTAAGGGTGGCGTGAGACCTACCTGCTAGTAGTGAAGAGGGGTGACAGTGTCTACA 120  
DB 98 GCTAAGGGTGGCGTGAGACCTACCTGCTAGTAGTGAAGAGGGGTGACAGTGTCTACA 157  
QY 121 TCCTTTTACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACCTGGAATTCGTC 180  
DB 158 TCCTTTTACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACCTGGAATTCGTC 217  
QY 181 TTCCTCCGTACATCTCGACCTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACTGG 240  
DB 218 TTCCTCCGTACATCTCGACCTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACTGG 277  
QY 241 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACCTTTCTGCGAGGG 300  
DB 278 TTCACCTCTCGAGCCCTGCTACGACTGTGCCGACATGTGCCGACCTTTCTGCGAGGG 337  
QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
DB 338 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 397  
QY 361 GCTGAGCCCGAGGGCTGGCGGGCTGCACCGCGCGGGTGCAGATAGCCATCATGACC 420  
DB 398 GCTGAGCCCGAGGGCTGGCGGGCTGCACCGCGCGGGTGCAGATAGCCATCATGACC 457  
QY 421 TTCAAGAT 429  
DB 458 TTCAAGAGT 466

RESULT 11  
AJ446140  
LOCUS AJ446140 riken1 Gallus gallus cDNA clone 12j9rl, mRNA sequence.  
DEFINITION  
ACCESSION AJ446140  
VERSION AJ446140.1 GI:20213361

```
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 757)
AUTHORS Buerstedde J.M.
TITLE Gallus gallus bursal lymphocyte EST
JOURNAL Unpublished (2002)
COMMENT Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES
source
Location/Qualifiers
1..757
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="12j9r1"
/clone_lib="rikenl"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/note="CB inbred strain"
BASE COUNT 180 a 205 c 184 g 188 t
ORIGIN
Query Match 68.9%; Score 411.4; DB 9; Length 757;
Best Local Similarity 80.6%; Pred. No. 1.1e-112;
Matches 481; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
Qy 1 ATGCACAGCCTCTTGATGACCGGAGGAGTTCCTTTACCAATTCAAAATGTCGCTGG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 97 ATGCACAGCCTCTTGATGACCGGAGGAGTTCCTTTACCAATTCAAAATGTCGCTGG 156
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 GCTAAGGCGCGCTGAGACCTACCTGTGTACGTAGTGAAGAGCGCTGACAGTGTACA 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 157 GCCAAGGCGCTGAGAACCTACCTCTGTATGTTGTGAAGCGCGTGAAGTGTCTACA 216
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 121 TCCTTTTACCTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCCAGTGGTCTC 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 217 TCATGCTCCCTGGACTTTGGATACCTGCGTAACAAGATGGTTGCCATGTGGAGTCTC 276
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGTGCCGACATGTGCCGACTCACCTGG 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 277 TTCCTAGCTACATCTCAGCTTGGGACCTGGACCGGCGCTGTACCGCATCACATGG 336
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 241 TTCACCTCTGGAGCCCTCTACGACTGTGCCGACATGTGCCGACTTCTTCGCGAGGG 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 337 TTCACCTCTGGAGCCCTCTTATGACTGTGCCGACATGTGGCTGACTTCCTTCGTGCC 396
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 397 TACCCAAACTTGACCTCGGCATTTTACCTGCGCGCTCTACTTCTGTGAAGTCCGATGACT 516
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 421 TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAAACCATGAAAGAACTTTCAA 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 517 TTCAAAGATTCTTCTACTGCTGGACACGTTTGTGGAGAACAGGAAAGACATTCAA 576
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 481 GCTGGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGGCGATCCCTT 540
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 577 GCTGGGAAGGCTGCATGAAACTCTGTCCATCTGTCCAGGAACTCCGACGATCCCTT 636
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 541 TTGCCCTGTATGAGTTGACTTACGAGACGCAATTTTCGTAATTTGGACTTTGA 597
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 637 CTGCCACTGTATGAAGTAGATGATTACGAGATGCCTTTAAACTCTGGGACTTTGA 693
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
RESULT 12
```

```
AJ453647
LOCUS AJ453647 696 bp mRNA linear EST 22-APR-2002
DEFINITION rikenl Gallus gallus cDNA clone 35b23r1, mRNA sequence.
ACCESSION AJ453647
VERSION AJ453647.1 GI:20263743
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 696)
AUTHORS Buerstedde J.M.
TITLE Gallus gallus bursal lymphocyte EST
JOURNAL Unpublished (2002)
COMMENT Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
FEATURES
source
Location/Qualifiers
1..696
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="35b23r1"
/clone_lib="rikenl"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/note="CB inbred strain"
BASE COUNT 160 a 193 c 173 g 170 t
ORIGIN
```

```
Query Match 68.6%; Score 409.8; DB 9; Length 696;
Best Local Similarity 80.4%; Pred. No. 3.2e-112;
Matches 480; Conservative 0; Mismatches 117; Indels 0; Gaps 0;
Qy 1 ATGCACAGCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAATGTCGCTGG 60
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 97 ATGCACAGCCTCTTGATGAAGAGGAGTTCCTCTCTACAATTTCAAGAACCTGCGCTGG 156
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 GCTAAGGCTCGGCTGAGACCTACCTGTGCTAGTAGTGAAGAGGCGTGACAGTGTCTACA 120
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 157 GCCAAGGCGCGTGTGAACCTACCTCTGTATGTTGTGAAGCGCGTGCACAGTGTCTACA 216
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 121 TCCTTTTACCTGGACTTTGGTTATCTTCGCAATGAAGACGGCTGCCAGTGGAAATGCTC 180
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 217 TCATGCTCCCTGGACTTTGGATACCTGCGTAACAAGATGGTTGCCATGTGGAGTGTCTC 276
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGTGCCGCGCTGTACCGCGTCACTGG 240
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 277 TTCCTAGCTACATCTCAGCTTGGGACCTGGACCGGCGCTGTACCGCATCACATGG 336
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 241 TTCACCTCTCGGAGCCCTCTACGACTGTGCCGACATGTGCCGACTTCTTCGCGAGGG 300
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 337 TTCACCTCTCGGAGCCCTCTTATGACTGTGCCGACATGTGGCTGACTTCCTTCGTGCC 396
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 397 TACCCAAACTTGACCTCGGCATTTTACCTGCGCGCTCTACTTCTGTGAAGTCCGATGACT 516
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 421 TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAAACCATGAAAGAACTTTCAA 480
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 517 TTCAAAGATTCTTCTACTGCTGGACACGTTTGTGGAGAACAGGAAAGACATTCAA 576
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 481 GCTGGGAAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGGCGATCCCTT 540
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 577 GCTGGGAAGGCTGCATGAAACTCTGTCCATCTGTCCAGGAACTCCGACGATCCCTT 636
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 541 TTGCCCTGTATGAGTTGACTTACGAGACGCAATTTTCGTAATTTGGGACTTTGA 597
```

Qy	394	GC	GGGGTGC	CAAAATAG	CCATCAT	CAGCACTT	CAAGATTAT	TTTTTACT	GCTGGA	TACTTTT	451
Db	311	GCT	GGGTGC	AGATCG	GGGATC	GACCTT	CAAGAC	TATTTTT	ACTGCT	GGAATAC	370
Qy	454	GT	AAAAAC	CAATGA	AAAGAACT	TTCA	AGCCTT	CGGA	GGCTGC	ATGAAATTC	513
Db	371	GT	AAAAAT	CGTGA	AAAGAACT	TTCA	AGCCTT	CGGA	GGCTGC	ATGAAATTC	430
Qy	514	CT	CTCCAG	ACAGCT	TCGCGC	GCATCCT	TTTTTG	CCCCCT	GTATG	AGGTTG	573
Db	431	CT	AACCA	GACAACT	TCGCGC	GCATCCT	TTTTTG	CCCCCT	GTATG	AGGTTG	490
Qy	574	GC	ATTTCG	TACTTT	GGGACT	TTGA	597				
Db	491	GC	ATTTCG	TATG	TGTTGG	ATTGTA	514				
RESULT 14											
LOCUS	BF975096	60245679	bf1	NIH_MGC_48	Homo sapiens	cdna clone IMAGE:4336722	5',	linear	EST 22-JAN-2001		
DEFINITION	mRNA sequence.										
ACCESSION	BF975096										
VERSION	BF975096.1	GI:12342311									
KEYWORDS	EST.										
SOURCE	human.										
ORGANISM	Homo sapiens										
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.										
AUTHORS	1 (bases 1 to 593)										
TITLE	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .										
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)										
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: coapbs-remail.nih.gov Tissue Procurement: Louis W. Staudt, M.D., Ph.D. cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LICM1209 row: n column: 19 High quality sequence stop: 692.										
FEATURES	Location/Qualifiers										
source	1..693	/organism="Homo sapiens" /db_xref="taxon:9606" /clone_image="IMAGE:4336722" /tissue_type="NIH_MGC_48" /lab_host="DH10B (phage-resistant)" /note="Organ: B-cells; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GCCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: This is a NIH_MGC Library."									
BASE COUNT	179 a	167 c	162 g	185 t							
ORIGIN											
Query Match	62.5%	Score	373;	DB	12;	Length	693;				
Best Local Similarity	100.0%;	Pred. No.	3.9e-101;								
Matches	373;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;		
Qy	225	CT	ACCGG	CTC	ACCTGG	GTTC	ACCTCCT	CTCG	AGCCCT	CTGTAC	284
Db	1	CT	ACCGG	CTC	ACCTGG	TTAC	CTCCTCG	AGCCCT	CTGTAC	AGCTGTG	60
Qy	285	CG	ACTTTT	CTCG	AGGGAA	ACCCCA	ACCTC	AGTCT	CTG	AGGATCT	344



```

Db      61  CGACTTTCTGGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGGCTCTACTT 120
Qy      345  CTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGGTGACCGCGCGGGGTGCA 404
Db      121  CTGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCGGTGACCGCGCGGGGTGCA 180
Qy      405  AATAGCCATCATGACCTTCAAGATTATTTTACTGCTGGAATACCTTTGTAGAAAACCA 464
Db      181  AATAGCCATCATGACCTTCAAGATTATTTTACTGCTGGAATACCTTTGTAGAAAACCA 240
Qy      465  TGAAGAACCTTCAAGCCCTGGGAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACA 524
Db      241  TGAAGAACCTTCAAGCCCTGGGAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACA 300
Qy      525  GCTTCGGCGCATCCTTTTGCCCTGTATGAGGTTGATGACTTACGAGACGCATTTGCTAC 584
Db      301  GCTTCGGCGCATCCTTTTGCCCTGTATGAGGTTGATGACTTACGAGACGCATTTGCTAC 360
Qy      585  TTTGGGACTTTGA 597
Db      361  TTTGGGACTTTGA 373

RESULT 15
AJ450296
LOCUS      729 bp mRNA linear EST 19-APR-2002
DEFINITION riken1 Gallus gallus cdna clone 24m22r1, mRNA sequence.
ACCESSION AJ450296
VERSION AJ450296.1 GI:20217517
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
            Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 729)
            Buerstedde,J.M.
            Gallus gallus bursal lymphocyte EST
            Unpublished (2002)
AUTHORS Contact: Buerstedde JM
JOURNAL Cellular Immunology
COMMENT Heinrich-Pette-Institute
            Martinistr. 52, 20251 Hamburg, Germany
            Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.

FEATURES             source
     1..729
     /organism="Gallus gallus"
     /db_xref="taxon:9031"
     /clone="24m22r1"
     /clone_lib="riken1"
     /cell_type="bursal lymphocyte"
     /dev_stage="2-3 weeks old"
     /note="CB inbred strain"

BASE COUNT  173 a 199 c 187 g 170 t
ORIGIN

Query Match      61.3%; Score 366.2; DB 9; Length 729;
Best Local Similarity 79.8%; Pred. No. 4.5e-99;
Matches 457; Conservative 0; Mismatches 113; Indels 3; Gaps 2;

Qy      1  ATGACACGCTCTTTGATGACCGGAGGAGTTCCTTTACCAATTTCAAAAATGTCGCGTGG 60
Db      157  ATGACACGCTCTTTGATGAGGAGGAGGCTCTCTCTACAAATTTCAAGAACCTCGCGCTGG 216
Qy      61  GCTAAGGCTGGCGTACCTCTGCTGCTAGTGAAGAGCGCTGACAGTGCCTACA 120
Db      217  GCCAAGCGCTGCTGAAACCTACCTCTGTATGTTGTAAGCGCCGCTGACAGTGCCTACA 276
Qy      121  TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGAACGCTGCCACAGTGGAAATTCGTC 180
Db      277  TCATGCTCCCTGGACTTTGGTATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 336
```

```

Qy      181  TTCTCTCGCTACATCTCTCGGACTGGGACCTAGACCCCTGGCCGCTGCTACCGGCTCACCTGG 240
Db      337  TTCTCTAGCTACATCTCAGCTTTGGAGCTGGAGCCAGCCGCTGCTACCGCATCACATGG 396
Qy      241  TTCAACCTCCCTGGAGCCCTGCTACGACTGTGCCCGACATGTGGCCGAGCTTTCTCGGAGGG 300
Db      397  TTCAACCTCCCTGGAGCCCTGTTATGACTGTGCCCGACATGTGGCTGACTTCTCTTGGTGCC 456
Qy      301  AACCCCAACCTCAGCTCTGAGGATCTTCACCGCGCGGCTCTACTTCTGTGAGGACCGCAAG 360
Db      457  TACCCAAACTTGACCCCTCCGCATTTTCACTGCCCGCCTCTACTTCTGTGAAGATCGCAAG 516
Qy      361  GCTGAGCCGAGGGGCTGCGGCGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
Db      517  GCTGAGCCTGAGGGGCTGAGACGCTGCACCGGCTGGGGCCCAATCGCCATCATGACT 576
Qy      421  TTCAAGATATTTTACTGCTGGAATACTTTTCTAGAAAACCATGAAAGAACCTTTCANA 480
Db      577  TTCAAGATATTTTCTTACTGCTGGAACACGTTTGTGGAGAACAGGGGAAAGACATTCANA 636
Qy      481  GCCTGGGAAGGCTGCATGAAAATTCAGTTTCAGTTTCAGTTTCAGTTTCAGTTTCAGTTTC 538
Db      637  GCCTGGGAAGGCTGCATGAAAATTCAGTTTCAGTTTCAGTTTCAGTTTCAGTTTCAGTTTC 696
Qy      539  TT-TTGGCCCTGTATGAGGTTGATGACTTACGA 570
Db      697  TTCTTGGCACTGTATGAAGTAGATGATTTACGA 729
```

Search completed: June 19, 2003, 11:07:28  
Job time : 808.004 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:01:09 ; Search time 25.0728 seconds  
(without alignments)  
7302.178 Million cell updates/sec

Title: US-09-966-880A-7\_COPY\_80\_676

Perfect score: 597  
Sequence: 1 atggacagcctcttgatga.....ttcgtacttgggactttga 597

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCtUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	114.8	19.2	610	1 US-08-816-241-2	Sequence 2, Appli
2	114.8	19.2	610	3 US-09-128-395-2	Sequence 2, Appli
3	63.6	10.7	891	1 US-08-687-895-2	Sequence 2, Appli
4	63.6	10.7	891	2 US-09-040-482-2	Sequence 2, Appli
5	60	10.1	879	1 US-08-158-682A-3	Sequence 3, Appli
6	56.4	9.4	879	1 US-08-158-682A-1	Sequence 1, Appli
7	56.4	9.4	879	1 US-08-015-203-1	Sequence 1, Appli
8	38	6.4	1929	4 US-09-380-420C-1	Sequence 1, Appli
9	37.8	6.3	922	4 US-09-189-060B-73	Sequence 1, Appli
10	36.6	6.1	4689	4 US-09-105-537-34	Sequence 34, Appl
11	36.6	6.1	36778	4 US-09-105-537-5	Sequence 5, Appli
12	36.6	6.1	38506	3 US-09-320-878-19	Sequence 19, Appl
13	36.4	6.1	7218	1 US-08-232-463-14	Sequence 14, Appl
14	35.4	5.9	2190	4 US-09-015-188-1	Sequence 1, Appli
15	35	5.9	80246	4 US-09-078-294-4	Sequence 4, Appli
16	35	5.9	80595	4 US-09-078-294-3	Sequence 3, Appli
17	34.6	5.8	1597	2 US-08-724-974A-1	Sequence 1, Appli
18	34.2	5.7	5159	3 US-08-146-930-3	Sequence 3, Appli
19	34.2	5.7	5159	3 US-08-458-240-3	Sequence 3, Appli
20	34.2	5.7	5159	5 PCF-US93-03993-3	Sequence 3, Appli
21	33.8	5.7	15872	4 US-09-105-537-1	Sequence 1, Appli
22	33.8	5.7	4403765	4 US-09-103-840A-2	Sequence 2, Appli
23	33.8	5.7	4411529	4 US-09-103-840A-1	Sequence 1, Appli
24	33.2	5.6	35081	2 US-08-752-760A-1	Sequence 1, Appli
25	33	5.5	4911	4 US-09-718-692-1	Sequence 1, Appli
26	33	5.5	4911	4 US-09-718-852-1	Sequence 1, Appli
27	33	5.5	4911	4 US-09-718-815-1	Sequence 1, Appli

28	32.8	5.5	36519	3 US-08-923-137-2	Sequence 2, Appli
29	32.6	5.5	1356	1 US-08-317-880-4	Sequence 4, Appli
30	32.6	5.5	1356	2 US-08-782-396-4	Sequence 4, Appli
31	32.6	5.5	3954	1 US-08-317-880-1	Sequence 1, Appli
32	32.6	5.5	3954	2 US-08-782-396-1	Sequence 1, Appli
33	32.2	5.4	796	4 US-09-280-116-107	Sequence 107, App
34	32.2	5.4	1943	4 US-09-282-305-7	Sequence 7, Appli
35	32	5.4	8878	1 US-08-759-444-2	Sequence 2, Appli
36	32	5.4	9880	3 US-08-680-897-1	Sequence 1, Appli
37	31.8	5.3	657	4 US-09-527-345-3	Sequence 3, Appli
38	31.4	5.3	1008	4 US-09-500-569-1	Sequence 1, Appli
39	31.4	5.3	1621	3 US-09-013-881-14	Sequence 14, Appl
40	31.4	5.3	4141	4 US-09-245-281-42	Sequence 42, Appl
41	31.4	5.3	4141	4 US-09-207-359B-42	Sequence 42, Appl
42	31.4	5.3	13842	4 US-09-105-537-30	Sequence 30, Appl
43	31.4	5.3	29629	4 US-09-729-995-3	Sequence 3, Appli
44	31.4	5.3	4411529	4 US-09-103-840A-1	Sequence 1, Appli
45	31.2	5.2	1392	3 US-08-957-063-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1  
US-08-816-241-2  
; Sequence 2, Application US/08816241  
; Patent No. 5804185  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/816,241  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0239 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 610 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: PROSTUT09  
; CLONE: 1646833  
US-08-816-241-2

Query Match 19.2%; Score 114.8; DB 1; Length 610;  
Best Local Similarity 53.9%; Pred. No. 1.9e-25;  
Matches 289; Conservative 0; Mismatches 232; Indels 15; Gaps 2;

[illegible]

154	ACGGCGCTCTCCGAAACCAAGGTGGATTCTGACACCCATTGTTCATSCAGAAAGGTGCTTC	213
Db		
184	CTCGCGTACATCTCGGACTGGGACCTAGACCCTGGCGCGTGTGTACCGCGTCACTGGTTC	243
QY		
214	CTCTCTGGTCTCGCAGCAGATACTGTCTCTCAACAAGTACCAGGTCACTGGTAC	273
Db		
244	ACCTCTGGAGGCCCTCGTAGGACTGTGCGCCGACATGTGCCGACATTTCTGCGAGGGAC	303
QY		

214	CTCTCTTTGGTTCTGCGACGACATACTGCTCTCCCTAACACAAAGTACCAGGTCACCTGGTAC	273
Db		
244	ACCTCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAGGGAAC	303
Qy		
274	ACATCTTGGAGCCCTTGGCCNAGCTGTGCAGGGAGGTGCCGAGTTCCTGCCAGGCAC	333
Db		
304	CCCAACCTCAGTCTGAGAGTCTTCCGCGCGGCTCTACTTCTGTGAGACGCCAAGGCT	363
Qy		
334	AGCAACGTGAATCTCACCATCTTCACGCGCGGCTCTACTACTTCCAGTATC---CATGT	390
Db		
364	GAGCCCGAGGGGCTGCGCGGCTGCACGCGCGGGGTGCAAAATAGCCATCATGACCTTC	423
Qy		
391	TACCAGGAGGGGCTCCGACGCTCAGTCAGGAAGGGGTGCGCTGTGGAGATCATGGACTAT	450
Db		
424	AAAGATATTTTTTACTGCTGGAATACATTTGTAGAAAACCATGAAGAAGCTTTCAAGCC	483
Qy		
451	GAGATATTTAAATATTTGTGGGAAACATTTGTGTACAAATGATGAGCCATTCAGACCT	510
Db		
484	TGGGAAGGGCTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGGCGCATCCT	539
Qy		
511	TGGAAGGATTTAAAACCACTTCCTGACTTCTGAAAAGAAGGTACGGGAGAGTCT	566
Db		

STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,395  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/816,241  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:

Db	606	GGTGCACCTTTCATGTGGGAGGAGCCGGAGATCCAG---GCTGCTCTGAAGAAGACTGAAG	662
QY	391	CGGCGCGGGTGCAAATAGCCATCATGACCTTCAAAGATTATTTTACTGCTGGATACT	450
Db	663	GAGGCTTGGCTGTAAACTGGGCATCATGAAGCCCGGAGACTTCGAATATGCTGGCAGAAT	722
QY	451	TTTGTGAGAAACCATCAAGAGACTTTTCAAAGCCTGGGAAGGGCTGCATGAAAT	504
Db	723	TTTGTGGAGCAAGAAGAGGTTGAATCCAAAGGCTTTCARCCCTGGGAGGACATT	776

RESULT 5  
US-08-158-682A-3  
Sequence 3, Application US/08158682A



CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/015,203  
FILING DATE: 19930209  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Coolley, Ronald B.  
REGISTRATION NUMBER: 27,187  
REFERENCE/DOCKET NUMBER: ARCD:069  
TELEPHONE: (312) 744-0090  
TELEFAX: (312) 245-4961  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 879 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-015-203-1

Query Match 9.4%; Score 56.4; DB 1; Length 879;  
Best Local Similarity 53.4%; Pred. No. 1.3e-07;  
Matches 142; Conservative 0; Mismatches 121; Indels 3; Gaps 1;

QY 197 CGGACTGGGACCTAGACCTGCGCGCTGCTACCGGCTACCTGGTTACCTCCTGGAGCC 256  
DB 255 CAGAAGATACCTTTGTCCAAACACCATGCTCCATTACCTGGTTCTGTCTGGAGTC 314  
QY 257 CTGCTACGACTGTGCGCGACATGTGGCGACTTTCTGGAGGGAACCCCAACCTCAGTC 316  
DB 315 CCGTGGGAGTGTCTCCAGGGCCATTACAGATTTTGGAGCGATACCCCATGTAATC 374  
QY 317 TGAGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAGGCTGAGCCGAGGGGC 376  
DB 375 TTTTATTATATAGCAGCGCTTTATCACACGACGATCCTCGAAATCGCC---AAGGAC 431  
QY 377 TCGGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACCTTCAAGATTATTTT 436  
DB 432 TCAGGAGCTTATTAGCAGCGGTGTACTATCCAGATCATGACGGAAGAGTCTGGCT 491  
QY 437 ACTGCTGGAATACTTTTGTAGAAAAC 462  
DB 492 ACTGCTGGAGGAATTTGTCAACTAC 517

RESULT 8  
US-09-380-420C-1  
Sequence 1, Application US/09380420C  
Patent No. 6300544

## GENERAL INFORMATION:

APPLICANT: Halkier, Barbara  
Kahn, Rachel  
Moller, Birger  
TITLE OF INVENTION: Cytochrome P450 Monooxygenases  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESS: Syngenta Patent Dept.  
STREET: 3054 Cornwallis Road  
CITY: RTP  
STATE: NC  
COUNTRY: USA  
ZIP: 27709  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/380,420C  
FILING DATE: 12-No. 6300544-1999  
CLASSIFICATION: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: S-21251A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1929 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
CLONE: P450ox  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 81..1673  
SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-380-420C-1

Query Match 6.4%; Score 38; DB 4; Length 1929;  
Best Local Similarity 53.3%; Pred. No. 0.072;  
Matches 80; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 236 CTGTTTACCTTCCTGGAGCCCTGCTAGACTGTCCCGACATGTGGCGGACTTTCTGC 295  
DB 793 CCAGCTTCTCCCGGAGGACTTCTCCCAACGCGCGCGCTCGCGGACCGCTCT 852  
QY 296 GAGGAAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACC 355  
DB 853 CGGGCTTCTCGCGCGCGGAGCGCATCTTCAACGAGGTCGACGCTCTTCTTCGAGAAG 912  
QY 356 GCAAGCTGAGCCGCGGCGGTGCGGCGGC 385  
DB 913 TCATCAGCACGACATGGACCGCGCGGC 942

## RESULT 9

US-09-189-060B-73  
Sequence 73, Application US/09189060B  
Patent No. 6270968

## GENERAL INFORMATION:

APPLICANT: Dalboge, Henrik  
APPLICANT: Sandal, Thomas  
APPLICANT: Kauppinen, Markus  
APPLICANT: Borge, Diderichsen  
TITLE OF INVENTION: Method Of Providing No. 6270968el DNA Sequences  
FILE REFERENCE: 4772.204-US  
CURRENT APPLICATION NUMBER: US/09/189,060B  
CURRENT FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: PCT/DK97/00216  
PRIOR FILING DATE: 1997-05-12  
NUMBER OF SEQ ID NOS: 74  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 73  
LENGTH: 922  
TYPE: DNA  
ORGANISM: Hybrid  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (2)...(922)  
US-09-189-060B-73

Query Match 6.3%; Score 37.8; DB 4; Length 922;

RESULT 11  
US-09-105-537-5  
; Sequence 5, Application US/09105537A  
; Patent No. 6265202  
; GENERAL INFORMATION:  
; APPLICANT: Sherman, D.H.  
; APPLICANT: Liu, H.

;  
 ; LENGTH: 38506  
 ;  
 ; TYPE: DNA  
 ;  
 ; ORGANISM: Streptomyces venezuelae  
 US-09-320-878-19

RESULT 12  
US-09-320-878-19  
: Sequence 19. Apollication US/09320878A

[illegible]

```

RESULT 14
US-09-015-188-1
; Sequence 1, Application US/09015188C
; Patent No. 6399358
; GENERAL INFORMATION:
; APPLICANT: Williams, Kevin J
; APPLICANT: Tabas, Ira
; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
; FILE OF INVENTION: 6-Sulfotransferase
; FILE REFERENCE: JEFF-0231
; CURRENT APPLICATION NUMBER: US/09/015,188C
; CURRENT FILING DATE: 1998-01-29
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2190
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-015-188-1

Query Match          5.9%; Score 35.4; DB 4; Length 2190;
Best Local Similarity 48.7%; Pred. No. 0.47;
Matches 96; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

Qy      162 CTGCCACGTGGGAATTGCTTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCTGGCGG 221
        ||| ||||| ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      343 CTCCTTCGTGGGCCAGCTCTTCAACCAAGCACCTGGACGTCCTTACCTGTTTGAGCCCCT 402

Qy      222 CTGCTACCGGCCTACCTGTTGCTGACCTTCCTGGAGCCCTCTGCTACGACTGTGCCCGACATGT 281
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      403 CTACCACGTCCTCAAGAACACGCTCATCCCGCTTCACCCAGGCAAGACCCGCGGACCG 462

Qy      282 GGCCGACTTCTCCGAGGAAACCCCAACCTCAGCTGAGGATCTTCACCGCGGCCCTCTA 341
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      463 GCGGGTCTAGCTATGGCCGACCGGACCTCTGCGGAGCCCTACGACTGCGACCTCTTA 522

Qy      342 CTTCTGTGAGGACCGCA 358
        |||| ||| ||| |||
Db      523 CTCCTGGAGAATACTACA 539

RESULT 15
US-09-078-294-4
; Sequence 4, Application US/09078294
; Patent No. 6265211
; GENERAL INFORMATION:
; APPLICANT: Choo, Kong-Hong Andy
```

```

; APPLICANT: Du Sart, Desiree
; APPLICANT: Cancilla, Michael R.
; TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE
; FILE REFERENCE: Davies Col
; CURRENT APPLICATION NUMBER: US/09/078,294
; CURRENT FILING DATE: 1998-05-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 80246
; TYPE: DNA
; ORGANISM: Nucleotide sequence of NC-contig
US-09-078-294-4

Query Match          5.9%; Score 35; DB 4; Length 80246;
Best Local Similarity 54.2%; Pred. No. 3.5;
Matches 71; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

QY      421 TTCAAGATTATTTTACTGCTGGAATACTTTGTAGAAAAACCATGAAGAACCTTTCAA 480
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      67319 TTCAAAATTGTTTCTTCTCCAAAGTAGCATATAGTCAATAATCTTAAAGAGAAGCAAA 67378
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      481 GCCTGGGAAGGCGTCATCGAATAATTCAGTTCGTCTCTCCAGACAGCTTTCGCGCATCCTT 540
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      67379 GAAGGGGAAGCACTGAACCAAAATTGCTCTTTTGTACCTGCCTCAAGATGCAGAGTT 67438
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY      541 TTGCCCCCTGTA 551
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      67439 CTCTACCTGGA 67449
        ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: June 19, 2003, 11:13:56  
Job time : 31.0728 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:06:39 ; Search time 80.0992 Seconds  
(without alignments)  
10937.110 Million cell updates/sec

Title: US-09-966-880A-7\_COPY\_80\_676  
Perfect score: 597  
Sequence: 1 atggacagcctcttgatga.....ttcacttgggactttga 597

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1042519 seqs, 733713590 residues  
Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_NA.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	597	100.0	2818	9	US-09-966-880A-7
2	475.4	79.6	2440	9	US-09-966-880A-1
3	273.4	45.8	6564	9	US-09-966-880A-10
4	273.4	45.8	11204	9	US-09-966-880A-35
5	271	45.4	271	9	US-09-966-880A-13
6	148	24.8	148	9	US-09-966-880A-12
7	118.4	19.8	1534	10	US-09-729-674-173
8	114.8	19.2	987	10	US-09-925-300-699
9	114.8	19.2	1143	9	US-10-098-841-268
10	114.4	19.2	116	9	US-09-966-880A-14
11	110.4	18.5	1348	10	US-09-880-107-3282
12	104.4	17.5	2151	10	US-09-822-830A-359
13	78.4	13.0	371	9	US-09-918-995-30237
14	71.6	12.0	1404	9	US-10-198-846-13472
15	70.2	11.8	390	9	US-09-796-692-9051
16	70.2	11.8	390	9	US-10-040-862-9051
17	67	11.2	572	10	US-09-864-761-9416
18	66.6	11.2	328	9	US-09-796-692-5737
19	66.6	11.2	328	9	US-10-040-862-5737

20	63	10.6	476	9	US-09-918-995-32385	Sequence 32385, A
21	60	10.1	879	9	US-10-157-031-13	Sequence 13, Appl
22	54	9.0	2172	9	US-09-966-880A-15	Sequence 15, Appl
23	49.4	8.3	556	9	US-09-796-692-5061	Sequence 5061, Ap
24	49.4	8.3	556	9	US-10-040-862-5061	Sequence 5061, Ap
25	48	8.0	553	9	US-09-796-692-4504	Sequence 4504, Ap
26	48	8.0	553	9	US-10-040-862-4504	Sequence 4504, Ap
27	43.8	7.3	466	9	US-09-796-692-4007	Sequence 4007, Ap
28	43.8	7.3	466	9	US-10-040-862-4007	Sequence 4007, Ap
29	43.8	7.3	505	9	US-09-796-692-4443	Sequence 4443, Ap
30	43.8	7.3	505	9	US-10-040-862-4443	Sequence 4443, Ap
31	43.6	7.3	254	10	US-09-867-701-3576	Sequence 3576, Ap
32	43	7.2	997	10	US-09-864-761-9278	Sequence 9278, Ap
33	40.8	6.8	154	10	US-09-864-761-25835	Sequence 25835, A
34	38.4	6.4	261	9	US-10-015-219-763	Sequence 763, App
35	38.4	6.4	261	10	US-09-777-564-763	Sequence 763, App
36	38	6.4	1929	9	US-09-899-642-1	Sequence 1, Appl
37	37.2	6.2	473	9	US-10-046-938-1443	Sequence 1443, Ap
38	37.2	6.2	473	9	US-09-878-178-1443	Sequence 1443, Ap
39	37.2	6.2	473	9	US-10-146-502-1443	Sequence 1443, Ap
40	37	6.2	520	9	US-10-184-644-332	Sequence 332, App
41	37	6.2	520	9	US-10-184-634-332	Sequence 332, App
42	36.6	6.1	4689	9	US-09-860-846-34	Sequence 34, Appl
43	36.6	6.1	4689	9	US-09-988-384B-34	Sequence 34, Appl
44	36.6	6.1	4689	9	US-09-836-821-34	Sequence 34, Appl
45	36.6	6.1	4689	10	US-09-861-289-34	Sequence 34, Appl

ALIGNMENTS

RESULT 1  
US-09-966-880A-7  
; Sequence 7, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966,880A  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 2818  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (80)...(673)  
; FEATURE:  
; NAME/KEY: 5'UTR  
; LOCATION: (1)...(79)  
; FEATURE:  
; NAME/KEY: 3'UTR  
; LOCATION: (677)...(2818)  
US-09-966-880A-7

Query Match 100.0%; Score 597; DB 9; Length 2818;  
Best Local Similarity 100.0%; Pred. No. 6.1e-189;  
Matches 597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCCTCGG 60  
|||||

```

Db 80 ATGACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCACAAATGTCGGCTGG 139
Qy 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGTAGTGAAGAGCGGTGACAGTGCTACA 120
Db 140 GCTAAGGGTCGGCGTGAGACCTACCTGTGTAGTGAAGAGCGGTGACAGTGCTACA 199
Qy 121 TCCTTTTCACTGACCTTTGGTTATCTCGCAATAAAGAACGGCTGCCACGTGGATGCTC 180
Db 200 TCCTTTTCACTGACCTTTGGTTATCTCGCAATAAAGAACGGCTGCCACGTGGATGCTC 259
Qy 181 TTCTCTCGGTACATCTCGGACTCGGACCTAGACCTGGCGGCTGCTACCGGCTCACCTGG 240
Db 260 TTCTCTCGGTACATCTCGGACTCGGACCTAGACCTGGCGGCTGCTACCGGCTCACCTGG 319
Qy 241 TTCACCTCTCGGAGCCCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTGCGAGGG 300
Db 320 TTCACCTCTCGGAGCCCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTGCGAGGG 379
Qy 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGGCTCTACCTGTGTGAGGACCGCAAG 360
Db 380 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGGCTCTACCTGTGTGAGGACCGCAAG 439
Qy 361 GCTGAGCCCGAGGGCTGCGGCGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 420
Db 440 GCTGAGCCCGAGGGCTGCGGCGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 499
Qy 421 TTCAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAAGAACTTTCAA 480
Db 500 TTCAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAAGAACTTTCAA 559
Qy 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 560 GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGGCGCATCCTT 619
Qy 541 TTGCCCTCTGTATGAGTGTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTGA 597
Db 620 TTGCCCTCTGTATGAGTGTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTGA 676

```

```

RESULT 2
US-09-966-880A-1
; Sequence 1, Application US/09966880A
; Patent No. US2002016473A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/Jp00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2440
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (93)...(686)
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)...(92)
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (690)...(2440)

```

FEATURE:

```

; NAME/KEY: misc_feature
; LOCATION: (1)...(2440)
; OTHER INFORMATION: n = A,T,C or G
US-09-966-880A-1

```

```

Query Match 79.6%; Score 475.4; DB 9; Length 2440;
Best Local Similarity 87.3%; Pred. No. 2.8e-148;
Matches 521; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

```

```

Qy 1 ATGACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCACAAATGTCGGCTGG 60
Db 93 ATGACAGCCTCTTGATGACCGGAGGAAGTTCTTTACCAATTCACAAATGTCGGCTGG 152
Qy 61 GCTAAGGGTCGGCGTGAGACCTACCTGTGTAGTGAAGAGCGGTGACAGTGCTACA 120
Db 153 GCTAAGGGTCGGCGTGAGACCTACCTGTGTAGTGAAGAGCGGTGACAGTGCTACA 212
Qy 121 TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAAAGAACGGCTGCCACGTGGATGCTC 180
Db 213 TCCTTTTCACTGACCTTTGGTTATCTTCGCAATAAAGAACGGCTGCCACGTGGATGCTC 272
Qy 181 TTCTCTCGGTACATCTCGGACTCGGACCTAGACCTGGCGGCTGCTACCGGCTCACCTGG 240
Db 273 TTCTCTCGGTACATCTCGGACTCGGACCTAGACCTGGCGGCTGCTACCGGCTCACCTGG 332
Qy 241 TTCACCTCTCGGAGCCCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTGCGAGGG 300
Db 333 TTCACCTCTCGGAGCCCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTGCGAGGG 392
Qy 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGGCTCTACCTGTGTGAGGACCGCAAG 360
Db 393 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGGCTCTACCTGTGTGAGGACCGCAAG 452
Qy 361 GCTGAGCCCGAGGGCTGCGGCGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 420
Db 453 GCTGAGCCCGAGGGCTGCGGCGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 512
Qy 421 TTCAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAAGAACTTTCAA 480
Db 513 TTCAAAGATTATTTTACTGCTGGAATACCTTTTGTAGAAACCATGAAAGAACTTTCAA 572
Qy 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 573 GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGGCGCATCCTT 632
Qy 541 TTGCCCTCTGTATGAGTGTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTGA 597
Db 633 TTGCCCTCTGTATGAGTGTGATGACTTACGAGACGCAATTCGTACTTTGGGACTTTGA 689

```

RESULT 3

```

US-09-966-880A-10
; Sequence 10, Application US/09966880A
; Patent No. US2002016473A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/Jp00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10

```

```
; LENGTH: 6564
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-10

Query Match      45.8%; Score 273.4; DB 9; Length 6564;
Best Local Similarity 99.6%; Pred. No. 2.1e-80;
Matches 274; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 155 AGAACGGCTGCCACGTGGGAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACC 214
      |||
Db 2590 AGAACGGCTGCCACGTGGGAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACC 2649
      |||
QY 215 CTGGCGCTGCTACCGCGTCACCTGCTTCACTCTCGAGCCCTCTAGACTGTGCC 274
      |||
Db 2650 CTGGCGCTGCTACCGCGTCACCTGCTTCACTCTCGAGCCCTCTAGACTGTGCC 2709
      |||
QY 275 GACATGTGCCGACTTCTTCGAGGGAACCCCAACCTCAGTCTGAGATCTTCAACCGCGC 334
      |||
Db 2710 GACATGTGCCGACTTCTTCGAGGGAACCCCAACCTCAGTCTGAGATCTTCAACCGCGC 2769
      |||
QY 335 GCCTCTACTCTGTGAGGACCGAAGCTGAGCCCGGCTGCGCGGCTGCACCGCG 394
      |||
Db 2770 GCCTCTACTCTGTGAGGACCGAAGCTGAGCCCGGCTGCGCGGCTGCACCGCG 2829
      |||
QY 395 CCGGGTGCAATAGCCATCATGACCTTCAAAGAT 429
      |||
Db 2830 CCGGGTGCAATAGCCATCATGACCTTCAAAGGT 2864
      |||
```

## RESULT 4

```
US-09-966-880A-35
; Sequence 35, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR FILING DATE: 1999-12-27
; PRIOR FILING DATE: 1999-06-24
; PRIOR FILING DATE: 1999-03-29
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 11204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-35
```

```
Query Match      45.8%; Score 273.4; DB 9; Length 11204;
Best Local Similarity 99.6%; Pred. No. 2.9e-80;
Matches 274; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 155 AGAACGGCTGCCACGTGGGAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACC 214
      |||
Db 7805 AGAACGGCTGCCACGTGGGAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACC 7864
      |||
QY 215 CTGGCGCTGCTACCGCGTCACCTGCTTCACTCTCGAGCCCTCTAGACTGTGCC 274
      |||
Db 7865 CTGGCGCTGCTACCGCGTCACCTGCTTCACTCTCGAGCCCTCTAGACTGTGCC 7924
      |||
QY 275 GACATGTGCCGACTTCTTCGAGGGAACCCCAACCTCAGTCTGAGATCTTCAACCGCGC 334
      |||
Db 7925 GACATGTGCCGACTTCTTCGAGGGAACCCCAACCTCAGTCTGAGATCTTCAACCGCGC 7984
      |||
```

```
QY 335 GCCTCTACTCTGTGAGACCGCAAGCTGAGCCCGGCTGCGGGCTGCGGGCTGCGACCGC 394
      |||
Db 7985 GCCTCTACTCTGTGAGACCGCAAGCTGAGCCCGGCTGCGGGCTGCGGGCTGCGACCGC 8044
      |||
QY 395 CCGGGTGCAATAGCCATCATGACCTTCAAAGAT 429
      |||
Db 8045 CCGGGTGCAATAGCCATCATGACCTTCAAAGGT 8079
      |||

RESULT 5
US-09-966-880A-13
; Sequence 13, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR FILING DATE: 1999-12-27
; PRIOR FILING DATE: 1999-06-24
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 271
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-13
```

```
Query Match      45.4%; Score 271; DB 9; Length 271;
Best Local Similarity 100.0%; Pred. No. 2.2e-80;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 157 AACGGCTGCCACGTGGGAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACCT 216
      |||
Db 1 AACGGCTGCCACGTGGGAATGCTCTTCTCGCTACATCTCGGACTGGGACCTAGACCT 60
      |||
QY 217 GCGCGCTGCTACCGCGTCACCTGCTTCACTCTCGAGCCCTCTAGACTGTGCCGA 276
      |||
Db 61 GCGCGCTGCTACCGCGTCACCTGCTTCACTCTCGAGCCCTCTAGACTGTGCCGA 120
      |||
QY 277 CATGTGCCGACTTCTTCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCAACCGCGC 336
      |||
Db 121 CATGTGCCGACTTCTTCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCAACCGCGC 180
      |||
QY 337 CTCTACTCTGTGAGACCGCAAGCTGAGCCCGGCTGCGGGCTGCGGGCTGCGACCGCC 396
      |||
Db 181 CTCTACTCTGTGAGACCGCAAGCTGAGCCCGGCTGCGGGCTGCGGGCTGCGACCGCC 240
      |||
QY 397 GGGGTGCAATAGCCATCATGACCTTCAAAG 427
      |||
Db 241 GGGGTGCAATAGCCATCATGACCTTCAAAG 271
      |||
```

## RESULT 6

```
US-09-966-880A-12
; Sequence 12, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966, 880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
```

	Query Match	Best Local Similarity	Score 118.4;	DB 10;	Length 1534;
	Matches 223;	Conservative 0;	Mismatches 151;	Indels 6;	Gaps 1;
QY	158	ACGGTGTCCAGCTGGAAATGCTCTTCTCGGTACATCTCGGACTGGGACCTAGACCTG	217		
Db	811	RAGCGCGCATGCAAGACTGTGTTCTCTGGAGCTGATTCCTCTTTTGGAGCTGGACCTGG	870		

Query Match	19.2%;	Score 114.8;	DB 10;	Length 987;
Best Local Similarity	53.9%;	Pred. NO. 8.7e-28;		
Matches 289; Conservative	0;	Mismatches 232;	Indels 15;	Gaps 2;
QY	16	ATGAACGGAGAGTTCCTTTACCAATTCAAAATGTCGCCTGGCTAAGGCTGGCGGT	75	
Db	132	ATGATATCCAGGCACATTCTACTTCCAATTTAAAAACCTATGGAAGCCACGATCGGAAC	191	
QY	76	GAGACCTACCTGTGCTACGTAGTCAAGAGCGCTCACAGTGCTACATCCTTTTCACTGGAC	135	
Db	192	GAAACTTGGCTGTGCTTCACCGTGGAGGTTAAAGCGCGCGCTCAGTTGCTCTCTGGAAG	251	
QY	136	TTTGGTTATCTTCGCAA-----TAAGAACGGCTGCCACGTGGGAATTCCTCTCTC	183	
Db	252	ACGGCGCTCTTCCGAAACACAGGTGGATTCTTGAGACCATTGTCTATGCAGAAAGTGGCTTC	311	
QY	184	CTCGCTACATCTCGGACTCGGACCTGAGACCCCTGGCGGCTGCTACCGCGTCACTGGTTC	243	
Db	312	CTCTCTTGGTCTCGCAGCAGACACTGTCTCTTAACACAAAGTACCAGGTCACCTGGTAC	371	
QY	244	ACCTCCTGGAGCCCTGCTACGACTGTGCCCGACATGTGCCCCACTTTCGCGAGGGAAC	303	
Db	372	ACATCTTGGAGCCCTTGCCACGACTGTGANGGGAGGTGGCCGAGTTCCTTGGCCGACAC	431	
QY	304	CCCAACCTCAGTCTGAGGATCTTCCACGGCGCGCTCTACTCTGTGAGGACCGCAAGGCT	363	

## RESULT 9

QY	428	ATTATTTTAC	TGCTGGAAT	TCTTTT	TAGAAAA	ACCATG	AAAGAAC	TTTCA	AAAGCCTGGG	487
Db	1	ATTATTTT	TCTGCTG	CAATCT	TTTT	TAGAAAA	ACCAGAA	GAAC	TTTCA	AAAGCCTGGG
										60
QY	488	AAGGGCTGC	ATGAAAT	TTCAGT	TCTCT	CCTCC	ACACAG	CTTCGG	GGGCAT	CCTTTTG
										543
Db	61	AAGGGCTGC	ATGAAAT	TTCAGT	TCTCT	CCTCC	ACACAG	CTTCGG	GGGCAT	CCTTTTG
										116

RESULT 11

US-09-880-107-3282

; Sequence 3282, Application US/09880107

; Patent No. US20020142981A1

; GENERAL INFORMATION:

; APPLICANT: Horne, Darci T.

```

; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3282
; LENGTH: 1348
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U03891
; NAME/KEY: unsure
; LOCATION: (1)..(1348)
; OTHER INFORMATION: n = a o r c o r g o r t
US-09-880-107-3282

```

Query Match 18.5%; Score 110.4; DB 10; Length 1348;  
Best Local Similarity 59.3%; Pred. No. 3.1e-26;  
Matches 230; Conservative 0; Mismatches 146; Indels 12; Gaps 2;

158	Qy	ACGGCTGCCACGTGGAATTTGCTCTTCCTCGCTACATCTCGGACTGGGACCTAGACCCTG	217
275	Db	ACGGCGCCATCGCGAGCTGGCTCTTGGACCTGGTTCTTCTTTGCAGTTGGACCGG	334
218	Qy	GCGCGTGTACTCCGCTACCTTGGTTTCACTCTCTGGAGCCCTCCTGCTAC-----GACTGTG	271
335	Db	CCGAGATCTACAGGCTCACTTGGTTTCATCTCTGGAGCCCTCTTCTCTCGGGGCTGTG	394
272	Qy	CCGCACATGTGGCCGACTTTCTCGCGAGGAACCCCAACTCACTCTGAGATCTTCACCG	331
395	Db	CCGGGAAGTGGCTGCTTCTTCAGAGAACACACACGTGACATGCGCATCTTCGCTG	454
332	Qy	CGCGCTCTACTTCTGTGAGGACGCAAGCTGAGCCGAGGGCTGCGCGCGCTGCACC	391
455	Db	CCCGCATCTA-----TGATTACCAACCCCTATATAAGGAGCGCTGCAATGCTCGGG	508
392	Qy	GCGCGGGTGCAATAGCCATCATGACCTTCAAGATTATTTTACTGCTGGAATACTT	451
509	Db	ATGCTGGGGCCCAAGCTTCCATCATGACCTACGATGAATTTAAGCACTGCTGGGACCT	568
452	Qy	TTGTAGAAAAACCATGAAGAATTTTCAAAGCCTGGGAAGGCTGCATGAAATTCAGTTC	511
569	Db	TTGTGACCACCGGATGTCCTTCCAGCCCTGGGATGGATAGATGACACGCCAAG	628
512	Qy	GTCTCTCAGACAGCTTCGGCGCATCT	539
629	Db	CCCTGAGTGGGAGGCTCGGGCCATTCT	656

```

RESULT 12
US-09-822-830A-359
; Sequence 359, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fecthel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakär
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES EN
; FILE REFERENCE: GIN 6402

```

```

: CURRENT APPLICATION NUMBER: US/09/822.830A
: CURRENT FILING DATE: 2001-03-29
: PRIOR APPLICATION NUMBER: 60/195,604
: PRIOR FILING DATE: 2000-04-06
: NUMBER OF SEQ ID NOS: 631
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 359
: LENGTH: 2151
: TYPE: DNA
: ORGANISM: Homo sapiens
US-09-822-830A-359

```

```
Query Match      17.5%; Score 104.4; DB 10; Length 2151;
Best Local Similarity 58.4%; Pred. No. 4.1e-24;
Matches 223; Conservative 0; Mismatches 151; Indels 8; Gaps 2;
```

Qy	158	ACGGCTGCCACGTGGAAATTGCTTCTCTCGCGTACATCTCGGACTGGAGCTAGACCTG	217
Db	1450	AAGCGCCCATGCAGAGCTGCTCTCTGACGTGATTCCTTTTGAAGCTGACCTGG	1509
Qy	218	GCCGCTGCTACCGCGTCACCTGGTTTCACTTCTCGAGCCCTGCTACGACTGTGCCGAC	277
Db	1510	ACCAGGACTACAGGGTTTACCTGCTTCACTTCTCGAGCCCTGCTTCACTGTGCCGAGG	1569
Qy	278	ATGTGCGCGACTTTTCTCGAGGGAACCCCAACCTCAGTCTGAGATCTTCACCGCGCGCC	337
Db	1570	AAATGGCTAAATTCATTTCAAAAACAACACGTGACCTGTGCATCTTCACGTCCCGCA	1629
Qy	338	TCTACTTCTGTGAGGACCGAAGGCTGAGCCCGAGGGCTGCGCGGCTGCACGCGCGG	397
Db	1630	TCCTA-----TGATGATCAAGGAAGATGTTCAGGAGGGCTGGCACCTTGGCCGAGGCTG	1683
Qy	398	GGGTGCAATAGCCATCATGACCTTCAAGAATATATTTTACTGCTGGAATACTTTTGTAG	457
Db	1684	GGGCCAAAATTTCAATTAATACATACAGTGAATTTAAGCACTGCTGACACCTTTTG--G	1741
Qy	458	AAAACCATGAAGAACCTTTCAAAGCCTGGAAAGGCTGCATGAAATTCAGTTGCTCTCT	517
Db	1742	ACCACAGGAGTGTCCCTTCCAGCCCTGGGATGGACTAGATGACCACGACCAAGACCTGA	1801
Qy	518	CCAGACAGCTTCGGCGCATCCT	539
Db	1802	GTGGAGGCTGCGGGCAATCT	1823

```

RESULT 13
US-09-918-995-30237
; Sequence 30237, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30237
; LENGTH: 371
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-30237

```

```

Query Match      13.1%; Score 78.4; DB 9; Length 371;
Best Local Similarity 64.1%; Pred. No. 7.6e-16;
Matches 118; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

```

Qy	158	ACGGCTGCCACGTGGAATTGCTCTTCTCCGCTACATCTCGGACTGGGACCTAGACCGCTG	217
Db	171	AAGCGCGCATCGACAGCTGTGCTCTCTCGGACGTGATTCCTTTTGAAGCTGGACCTGG	230







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 14, 2003, 18:09:48 ; Search time 24.5713 Seconds  
(without alignments)  
6475.074 Million cell updates/sec

Title: US-09-966-880A-7\_COPY\_80\_676

Perfect score: 1101

Sequence: 1 atggacagcctcttgatgaa.....ttcgactcttgaggacttga 597

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+\_n2p.model -DEV=xlp  
-Q-/cgn2\_1/USPTO.spool/US09966880/runat\_14062003\_175523\_10297/app\_query.fasta\_1.9493  
-DB=A\_Geneseq\_101002 -QMT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09966880 -CGN\_1\_1\_632 -runat\_14062003\_175523\_10297 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NGS\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1086	98.6	198	21	AAB24198	Human activation-1
2	1008	91.6	198	21	AAB24197	Mouse activation-1
3	390	35.4	384	20	AA42383	Amino acid sequenc
4	390	35.4	384	22	AAU39075	Human secreted pro
5	390	35.4	384	23	ABBS5784	Human polypeptide
6	388	35.2	384	21	AA484437	Amino acid sequenc
7	369.5	33.6	190	19	AAW77092	Human RNA editing
8	369.5	33.6	190	21	AA111973	Human RNA editing
9	369.5	33.6	190	22	AA38905	Human polypeptide
10	369.5	33.6	210	22	AAU23550	Novel human enzyme
11	369.5	33.6	221	22	AA40691	Human polypeptide
12	369.5	33.6	222	21	AA37061	Human prostate can
13	369.5	33.6	222	22	AAU23537	Novel human enzyme
14	336.5	30.6	268	23	AAE15256	Human RNA metaboli
15	323.5	29.4	272	22	AAU23799	Novel human enzyme
16	309.5	28.1	163	23	AB97879	Human secretory po
17	274	24.9	224	23	AAE16630	Human AAD45360 pro
18	274	24.9	224	23	AAE15900	Human APOBEC2 prot
19	271	24.6	222	19	AAW52131	A novel human mRNA
20	262.5	23.8	367	22	AAO13149	Human polypeptide
21	208.5	18.9	261	22	AAU30167	Novel human secret
22	204.5	18.6	169	22	AAU30166	Novel human secret
23	203	18.4	113	21	AAU00666	Human secreted pro
24	199	18.1	236	15	AA38705	Apo-B RNA editing
25	198	18.0	127	22	AA395163	Human protein sequ
26	193	17.5	229	15	AA38704	Apo-B RNA editing
27	103.5	9.4	94	21	AAU01186	Human secreted pro
28	102	9.3	720	22	AB62508	Drosophila melanog
29	99.5	9.0	925	23	AAO14246	Human presenilin e
30	97.5	8.9	298	18	AAW43397	Human imidazole
31	97.5	8.9	1588	23	AB09437	H. influenzae DXR
32	95	8.6	293	22	ABG14103	Novel human diagno
33	94.5	8.6	568	22	ABG66474	Murine cell growth
34	92.5	8.4	559	18	AAW43396	Human imidazole
35	92.5	8.4	2342	22	AB60584	Drosophila melanog
36	90	8.2	242	22	AAU62666	Propionibacterium
c 37	90	8.7	581	20	AA14543	Human lipolysis st
c 38	90	8.7	581	21	AA395824	Human lipolysis st
c 39	90	8.7	581	22	AA359903	Human leptin fragm
c 40	90	8.7	630	20	AA14542	Human lipolysis st
c 41	90	8.7	630	21	AA395823	Human lipolysis st
c 42	90	8.7	630	22	AA359902	Human leptin fragm
c 43	90	8.7	649	20	AA14541	Human lipolysis st
c 44	90	8.7	649	21	AA395822	Human lipolysis st
c 45	90	8.7	649	22	AA359901	Human leptin fragm

ALIGNMENTS\*

RESULT 1

AAB24198  
ID AAB24198 standard; Protein; 198 AA.

XX AAB24198;

XX 05-FEB-2001 (first entry)

XX Human activation-induced cytidine deaminase SEQ ID NO:8.

XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; anti-allergic;  
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW Iga nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;

KW auto immunodeficiency syndrome; IgG subclass selection disorder.  
 XX Homo sapiens.  
 OS WO200058480-A1.  
 PN 05-OCT-2000.  
 XX 28-MAR-2000; 2000WO-JP01918.  
 XX 29-MAR-1999; 99JP-0087192.  
 PR 24-JUN-1999; 99JP-0178999.  
 PR 27-DEC-1999; 99JP-0371382.  
 XX (NLSB ) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 XX Honjo T, Muramatsu M;  
 XX WPI: 2000-611715/58.  
 DR N-PSDB; AAC55312.  
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 XX Claim 1; Page 140-141; 174pp; Japanese.  
 PS The present sequence is human activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC has cytidine activity similar to APOBEC-1. AID has anti-allergic,  
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and  
 CC dermatological activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
 CC disease, IgA nephritis, gamma-globulinemia, atopic dermatitis, allergic  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
 CC disease, Digeorge disease, ataxia telangiectasia, common variable  
 CC immunodeficiency disorder, MHC (major histocompatibility class) class  
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
 CC encoding AID may be used for gene therapy and the antibodies to the AID  
 CC protein may be used for diagnosis and treatment of these disorders.  
 XX SQ Sequence 198 AA;

Alignment Scores:  
 Pred. No.: 8.56e-115 Length: 198  
 Score: 1086.00 Matches: 198  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 98.64% Indels: 0  
 DB: 21 Gaps: 0

US-09-966-880A-7\_COPY\_80\_676 (1-597) x AAB24198 (1-198)

QY 1 ATGCAGACCTCTGTATGACCGGAGGAGCTTTCTTTACCAATTCAAAAATGTCGCTGG 60  
 DB 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
 QY 61 GCTAAGGTCGGCGTCAGACCTACCTCTGCTAGCTAGTGAAGAGCGGTGACGTGTACA 120  
 DB 21 AlaLysGlyArgArgGlnThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
 QY 121 TCCTTTCTACTGGACTTTGGTTATCTTCGCAATAAGAACGCTGCCACGTCGGATTGCTC 180  
 DB 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
 QY 181 TTCTCCGCTACATCTCGGACTCGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 240  
 DB 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80

QY 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCCGACATGTGCCGCACTTTCTGCGAGGG 300  
 DB 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
 QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
 DB 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
 QY 361 GCTGAGCCCGAGGGCTGCGCGGCTGCACCCGCGCGGTGCAAAATAGCCATCATGACC 420  
 DB 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
 QY 421 TTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAATCTTCAAA 480  
 DB 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
 QY 481 GCCTGGGAAGGCTGCATCAAAATTCAGTTCTGCTCTCCAGACACCTTCGGCGCATCTT 540  
 DB 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
 QY 541 TTGCCCTCTGATGAGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTT 594  
 DB 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198

RESULT 2  
 AAB24197  
 ID AAB24197 standard; Protein; 198 AA.  
 AC AAB24197;  
 XX 05-FEB-2001 (first entry)  
 DT Mouse activation-induced cytidine deaminase SEQ ID NO:2.  
 DE Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; anti-allergic;  
 KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IgA nephritis; gamma-globulinemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease;  
 KW auto immunodeficiency syndrome; IgG subclass selection disorder.  
 XX Mus musculus.  
 OS WO200058480-A1.  
 PN 05-OCT-2000.  
 XX 28-MAR-2000; 2000WO-JP01918.  
 PR 29-MAR-1999; 99JP-0087192.  
 PR 24-JUN-1999; 99JP-0178999.  
 PR 27-DEC-1999; 99JP-0371382.  
 XX (NLSB ) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 XX Honjo T, Muramatsu M;  
 XX WPI: 2000-611715/58.  
 DR N-PSDB; AAC55307.  
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 XX Claim 1; Page 131-132; 174pp; Japanese.  
 PS The present sequence is mouse activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and

CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and  
 CC dermatologic activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
 CC disease, IGA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
 CC immunodeficiency disorder, MHC (major histocompatibility class  
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
 CC IgE disorder, and Igg subclass selection disorder. The DNA sequences  
 CC encoding AID may be used for gene therapy and the antibodies to the AID  
 CC protein may be used for diagnosis and treatment of these disorders.  
 XX  
 SQ Sequence 198 AA;

Alignment Scores:  
 Pred. No.: 6.61e-106 Length: 198  
 Score: 1008.00 Matches: 183  
 Percent Similarity: 95.94% Conservative: 6  
 Best Local Similarity: 92.89% Mismatches: 8  
 Query Match: 91.55% Indels: 0  
 DB: 21 Gaps: 0

US-09-966-880A-7\_COPY\_80\_676 (1-597) x AAB24197 (1-198)

QY 1 ATGGACAGCTCTGATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCTGG 60  
 DB 1 MetAspSerLeuLeuMetLysGlnLysLysPheLeuTyrHisPheLysAsnValArgTrp 20  
 QY 61 GCTAAGGCGCGGCTGAGACCTACCTGTCTAGTAGTGAAGGCGTGACAGTGTACA 120  
 DB 21 AlaLysGlyArgHisGluThrTyrLeuCysTyrValValLysArgArgSerAlaThr 40  
 QY 121 TCCTTTTACTGACTTGGTTATCTTCGCAATTAAGACGGCTGCCACGTGGAAATGCTC 180  
 DB 41 SerCysSerLeuAspPheGlyHisLeuArgAsnLysSerGlyCysHisValGluLeuLeu 60  
 QY 181 TTCCTCCGTACATCTCGACCTGGGACCTAGACCTGCGCGCTACCGCTCACCTGG 240  
 DB 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
 QY 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCCCGACATGTGGCCGACTTCTTCGCGAGG 300  
 DB 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaGluPheLeuArgTrp 100  
 QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACGCGCGGCTCTACTCTGTGTAGGACCGCAAG 360  
 DB 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
 QY 361 GCTGAGCCGAGGCGTGGCGGCTGCACCGCGCGGCTGCAATAGCCATCATGACC 420  
 DB 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleGlyIleMetThr 140  
 QY 421 TTCAAAGATATTTTACTGCTGGAATACTTTGTAGAAAACCATGAAGAACTTCAAA 480  
 DB 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnArgGluArgThrPheLys 160  
 QY 481 GCTTGGAGGCGTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGGCGCATCTT 540  
 DB 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuThrArgGlnLeuArgArgIleLeu 180  
 QY 541 TTGCCCCCTATCAGCTTGTATGACTTACGAGAGCGCATTTCTGACTTTGGGA 591  
 DB 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgMetLeuGly 197

RESULT 3  
 AAY42383  
 ID AAY42383 standard; Protein; 384 AA.  
 XX  
 AC AAY42383;  
 XX

DT 09-DEC-1999 (first entry)  
 XX Amino acid sequence of lp547\_4.  
 DE  
 XX secreted protein; cDNA library; clone; transmembrane protein;  
 KW signal sequence cloning; hybridization cloning; gene therapy;  
 KW receptor.  
 KW  
 XX Homo sapiens.  
 OS  
 FH Key Location/Qualifiers  
 PE Peptide /label= Leader/signal peptide  
 FT Protein 291..384  
 FT /label= Mature protein  
 XX  
 PN WO9942470-A1.  
 XX  
 XX 26-AUG-1999.  
 XX  
 PF 18-FEB-1999; 99WO-US03458.  
 XX  
 XX 18-FEB-1998; 98US-0075038.  
 PR  
 PR 17-FEB-1999; 99US-0251600.  
 XX  
 PA (GEMY ) GENETICS INST INC.  
 XX  
 XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Merberg D;  
 PI Treacy M, Agostino MJ, Steininger RJ;  
 XX  
 DR WPI: 1999-518580/43.  
 DR N-PSDB; AA220856.  
 XX  
 XX New polynucleotides encoding human secreted proteins used for  
 PT therapeutic, diagnostic and research purposes.  
 PT  
 XX  
 PS Claim 17; Page 104-105; 125pp; English.  
 XX  
 CC This is the amino acid sequence of the lp547\_4 protein, which is  
 CC derived from the lp547\_4 clone isolated from a human adult blood cDNA  
 CC library.  
 CC The PNs and proteins of the invention are predicted to have biological  
 CC activities which would make them suitable for treating, preventing or  
 CC ameliorating medical conditions in humans and animals, although no  
 CC supporting data is given. Suggested activities include nutritional  
 CC activity, cytokine and cell proliferation/differentiation activity,  
 CC immune stimulating (e.g. as vaccines) or suppressing activity,  
 CC hematopoiesis regulating activity, tissue growth activity,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic  
 CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
 CC activity, cadherin/tumor invasion suppressor activity, and tumor  
 CC inhibition activity. The PNs are also stated to be useful for gene  
 CC therapy. Other activities include inhibiting the growth, infection or  
 CC function of bacteria, fungi, viruses and other parasites; effecting  
 CC bodily characteristics such as, e.g. weight, color, skin, etc.,  
 CC effecting biorhythms or cardiac cycles; enhancing fertility; treatment  
 CC of depression; treatment of pain; hormonal or endocrine activity.  
 XX  
 SQ Sequence 384 AA;

Alignment Scores:  
 Pred. No.: 2.22e-35 Length: 384  
 Score: 390.00 Matches: 83  
 Percent Similarity: 61.62% Conservative: 31  
 Best Local Similarity: 44.86% Mismatches: 59  
 Query Match: 35.42% Indels: 12  
 DB: 20 Gaps: 3

US-09-966-880A-7\_COPY\_80\_676 (1-597) x AAY42383 (1-384)

QY 16 ATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCGCTGGTACGGTGGCGGT 75  
 DB 197 MetAspProThrPheThrPheAsnPheAsnGluProTrpValArgGlyArgHis 216

QY 76 GAGACCTACTGTGCTACCTAGTGAAGAGCGCTGACAGTGTCTACATCTCTTTCACCTGGAC 135  
 DB 217 GluThrTyrLeuCysTyrGluValGluArgMetHisAsnAspThrTrpValLeuLeuAsn 236  
 QY 136 -----TTTGGTTATCTTCGCATTAAG-----AACGGCTGC 165  
 DB 237 GlnArgArgGlyPheLeuLeuAsnGlnAlaProHisLysHisGlyPheLeuGluGlyArg 256  
 QY 166 CACGTGGGAATGCTCTCTCCGCTACATCTCGGACTGGAGCTAGACCTGGCGCTGC 225  
 DB 257 HisAlaGluLeuCysPheLeuAspValIleProPheTrpLysLeuAspLeuAspGlnAsp 276  
 QY 226 TACCGCGTCACCTGGTTCACCTCCCTGGAGCCCTGCTAGCAGTGTGCTGCTGCTGCTG 285  
 DB 277 TyrArgValThrCysPheThrSerTrpSerProCysPheSerCysAlaGlnGluMetAla 296  
 QY 286 GACTTCTTCGCGAGGAGCAACCTCAGTCTGAGATCTTCACCGCGCTGCTGCTGCTGCTG 345  
 DB 297 LysPheIleSerLysAsnLysHisValSerLeuLeuCysIlePheThrAlaArgIleTyr--- 315  
 QY 346 TGTGAGGACCGCAAGCTGAGCCGCGGCTGCGCGCTGCACCGCGCGGTGCAA 405  
 DB 316 ---AspAspGlnGlyArgCysGlnGluGlyLeuArgThrLeuAlaGluAlaGlyAlaLys 334  
 QY 406 ATAGCCATCATGACCTTCAAGATTATTTTACTGCTGGATATCTTTGTAGAAAACCAT 465  
 DB 335 IleSerIleMetThrTyrSerGluPheLysHisCysTrpAspThrPheValAspHisGln 354  
 QY 466 GAAAGAACTTCAAGCGCTGGAAGGCTGCATGAAATTCAGTTCGCTCTCAGACAG 525  
 DB 355 GlyCysProPheGlnProTrpAspGlyLeuAspGluHisSerGlnAspLeuSerGlyArg 374  
 QY 526 CTTCGGCGCATCTT 540  
 DB 375 LeuArgAlaIleLeu 379

RESULT 4

ID AAU39075 standard; Protein; 384 AA.  
 XX AAU39075;  
 AC AAU39075;  
 XX 16-JAN-2002 (first entry)  
 DT Human secreted protein lp547\_4.  
 DE Human; secreted protein; antiinflammatory; immunosuppressive;  
 KW neurotropic; neuroprotective; antiarthritic; antimicrobial; vulnary;  
 KW cytosatic; antidiabetic; virucide; antifertility; anticonvulsant;  
 KW vasotropic; antiparkinsonian; immunostimulant; dermatological;  
 KW antirheumatic; antitumor; antiulcer; osteopathic; tranquiliser;  
 KW cerebroprotective; cytokine; cell proliferation; cell differentiation;  
 KW immune deficiency; severe combined immunodeficiency; SCID; tumour;  
 KW autoimmune disorder; multiple sclerosis; rheumatoid arthritis;  
 KW graft-versus-host disease; myeloid deficiency; wound healing; ulcer;  
 KW periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; infection; cardiac disease;  
 KW stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;  
 KW food supplement; vaccine.  
 XX Homo sapiens.  
 OS  
 XX W020175068-A2.  
 PN  
 XX 11-OCT-2001.  
 PD  
 XX 22-MAR-2001; 2001WO-US09369.  
 PF  
 XX 30-MAR-2000; 2000US-0539330.  
 PR  
 XX 04-DEC-2000; 2000US-0729674.  
 XX  
 PA (GEM ) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;  
 PI Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;  
 PI Clark H, Fechtel K, Merberg D;  
 XX WPI; 2001-639363/73.  
 DR N-PSDB; AAS59293.  
 XX  
 PT Secreted human proteins, useful as vaccine for treating various  
 PT diseases such as autoimmune disorders (e.g. multiple sclerosis), and  
 PT nervous system disorders (e.g. stroke) -  
 PS Disclosure; Page 580-581; 619pp; English.  
 XX  
 CC The invention relates to novel human secreted proteins, the nucleic  
 CC acids encoding them. The protein may exhibit cytokine, cell proliferation  
 CC or cell differentiation activity or may induce production of other  
 CC cytokines in certain cell populations and may exhibit immune stimulating  
 CC or immune suppressing activity, which is useful for the treatment of  
 CC various immune deficiencies and disorders e.g. severe combined  
 CC immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,  
 CC systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary  
 CC inflammation. The proteins are also useful in the treatment of diseases  
 CC and disorders including tissue, skin and organ transplantation and in  
 CC graft-versus-host diseases (GVHD). In the induction of tumour immunity,  
 CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,  
 CC in the treatment of burns, incisions and ulcers; as well as in treatment  
 CC of periodontal disease, osteoporosis or osteoarthritis, mediated by  
 CC inflammatory processes, diseases of the peripheral nervous system,  
 CC Alzheimer's, Parkinson's disease, Huntington's disease,  
 CC amyotrophic lateral sclerosis, and Shy-Drager syndrome, infections,  
 CC infarction of cardiac and central nervous system vessel e.g. stroke,  
 CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The  
 CC protein, having activin- or inhibin-related activities is useful as a  
 CC contraceptive based on the ability of inhibins to decrease fertility in  
 CC female mammals and decrease spermatogenesis in male mammals. The  
 CC proteins and nucleic acids are also useful as food supplements. The  
 CC present sequence represents a secreted protein of the invention.  
 XX  
 SQ Sequence 384 AA;

Alignment Scores:  
 Pred. No.: 2.22e-35 Length: 384  
 Score: 390.00 Matches: 83  
 Percent Similarity: 61.62% Conservative: 31  
 Best Local Similarity: 44.86% Mismatches: 59  
 Query Match: 35.42% Indels: 12  
 DB: 22 Gaps: 3  
 US-09-966-880A-7\_COPY\_80\_676 (1-597) x AAU39075 (1-384)  
 QY 16 ATGAACGGAGGAGTTCCTTTTACCAATTCACAAATTCGCGTGGCTAGGCTCGCGGT 75  
 DB 197 MetAspProProThrPheThrPheAsnPheAsnAsnGluProTrpValArgGlyArgHis 216  
 QY 76 GAGACCTACTGTGCTACGTAGTGAAGAGCGCTGACAGTGTCTTTCACCTGGAC 135  
 DB 217 GluThrTyrLeuCysTyrGluValGluArgMetHisAsnAspThrTrpValLeuLeuAsn 236  
 QY 136 -----TTTGGTTATCTTCGCATTAAG-----AACGGCTGC 165  
 DB 237 GlnArgArgGlyPheLeuLeuAsnGlnAlaProHisLysHisGlyPheLeuGluGlyArg 256  
 QY 166 CACGTGGGAATGCTCTCTCCGCTACATCTCGGACTGGAGCTAGACCTGGCGCTGC 225  
 DB 257 HisAlaGluLeuCysPheLeuAspValIleProPheTrpLysLeuAspLeuAspGlnAsp 276  
 QY 226 TACCGCGTCACCTGGTTCACCTCCCTGGAGCCCTGCTAGCAGTGTGCTGCTGCTGCTG 285  
 DB 277 TyrArgValThrCysPheThrSerTrpSerProCysPheSerCysAlaGlnGluMetAla 296  
 QY 286 GACTTCTTCGCGAGGAGCAACCTCAGTCTGAGATCTTCCACCGCGCTCTACTTTC 345

Db 297 LysPheIleSerLysAsnLysHisValSerLeuCysIlePheThrAlaArgIleTyr--- 315  
QY 346 TGTGAGGACCGCAAGGCTGAGCCGAGGGGCTGCGGCTGCACCGCGCGGGTGCAC 405  
Db 316 ---AspAspGlnGlyArgCysGlnGluGlyLeuArgThrLeuAlaGluAlaLys 334  
QY 406 ATAGCCATCATGCTTCAAGATTAATTTTACTGCTGGAATACTTTGTAGAAACCAT 465  
Db 335 IleSerIleMetThrTyrSerGluPheLysHisCysTrpAspThrPheValAspHisGln 354  
QY 466 GAAGAAGACTTCAAGCCCTGGGAGGCTGCATGAAATTCAGTTCTCTCCACACAG 525  
Db 355 GlyCysProPheGlnProTrpAspGlyLeuAspGluHisSerGlnAspLeuSerGlyArg 374  
QY 526 CTTCCGGCGCATCCTT 540  
Db 375 LeuArgAlaIleLeu 379  
RESULT 5  
ABB55784  
ID ABB55784 standard; Protein; 384 AA.  
AC ABB55784;  
XX  
XX 14-FEB-2002 (first entry)  
XX Human polypeptide SEQ ID NO 174.  
XX Human; clone bd306-7; clone yb8-1; ATCC number 98599; gene therapy;  
KW immune disorder; bacterial infection; fungal infection; cancer; tumour;  
KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;  
KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;  
KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;  
KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnery;  
KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;  
KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulator;  
KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory.  
XX Homo sapiens.  
XX OS  
XX US2001039335-A1.  
XX  
XX 08-NOV-2001.  
XX  
XX 04-DEC-2000; 2000US-0729674.  
XX  
XX 10-APR-1997; 97US-126425P.  
PR 04-DEC-1997; 97US-067454P.  
PR 02-DEC-1997; 97US-068379P.  
PR 02-JAN-1998; 98US-070346P.  
PR 07-JAN-1998; 98US-070643P.  
PR 08-JAN-1998; 98US-070755P.  
PR 13-JAN-1998; 98US-071304P.  
PR 22-JAN-1998; 98US-072134P.  
PR 30-JAN-1998; 98US-073095P.  
PR 18-FEB-1998; 98US-075038P.  
PR 30-MAR-2000; 2000US-0539330.  
PR 23-NOV-1998; 98US-0197886.  
XX  
XX (JACO/) JACOBS K.  
PA (MCCO/) MCCOY J M.  
PA (LAVA/) LAVALLIE E R.  
PA (COLL/) COLLINS-RACIE L A.  
PA (EVAN/) EVANS C.  
PA (MERE/) MERBERG D.  
PA (TREA/) TREACY M.  
PA (AGOS/) AGOSTINO M J.  
PA (STEL/) STEININGER R J.  
PA (SPAU/) SPAULDING V.  
PA (WONG/) WONG G G.  
PA (CLAR/) CLARK H. K.  
XX (FECH/) FECHTEL K.

PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;  
PI Wong GG, Clark H, Fechtel K;  
XX  
XX WPI: 2002-040725/05.  
XX N-PSDB; ABA90962.  
XX  
XX New secreted proteins and encoding polynucleotides, useful in gene  
PT therapies, particularly for preventing or treating autoimmune  
PT disorders, cancer, graft-versus-host disease, wound, osteoporosis,  
PT stroke or inflammations  
XX  
XX Disclosure; Page 311-312; 349pp; English.  
XX  
XX The invention relates to isolated polynucleotides (ABA90876-ABA90968 and  
CC ABA90980) and encoded proteins (ABB55698-ABB55800), especially  
CC polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90985) and  
CC proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in  
CC clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1  
CC are deposited with the American Type Culture Collection (ATCC) with  
CC accession number 98599. The polynucleotides and encoded polypeptides have  
CC cytostatic, anti-inflammatory, immunomodulator, vulnery,  
CC neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic  
CC and anti-inflammatory activity and acting as cytokine modulators,  
CC haematopoiesis regulators, tissue growth modulators and/or cadherin  
CC suppressors. The polypeptides and polynucleotides are useful in gene  
CC therapies, particularly for preventing, treating or ameliorating any of  
CC the following diseases: immune deficiency and disorders; e.g. bacterial  
CC or fungal infections, autoimmune disorders, cancer, systemic lupus  
CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell  
CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or  
CC osteoarthritis; central and peripheral nervous system diseases and  
CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's  
CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;  
CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis  
CC or systemic inflammatory response syndrome, ischaemia-reperfusion  
CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or  
CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus  
CC foliaceus.  
XX  
XX Sequence 384 AA;  
SQ  
Alignment Scores:  
Pred. No.: 2.22e-35 Length: 384  
Score: 390.00 Matches: 83  
Percent Similarity: 61.62% Conservative: 31  
Best Local Similarity: 44.86% Mismatches: 59  
Query Match: 35.42% Indels: 12  
DB: 23 Gaps: 3  
US-09-966-880A-7\_COPY\_80\_676 (1-597) x ABB55784 (1-384)  
QY 16 ATGACCGGAGGAGTTTCTTACCAATTCAAAATGTCGCTGGGCTAGGGTCGGGT 75  
Db 197 MetAspProThrPheThrPheAsnGluProTrpValArgGlyArgHis 216  
QY 76 GAGACCTACCTGCTACGTAGTGAAGAGCGGTGCACAGTGTCTCTTCTACCTGGAC 135  
Db 217 GluThrTyrLeuCysTyrGluValGluArgMethHisAsnAspThrTrpValLeuLeuAsn 236  
QY 136 -----TTTGGTTATCTTCGCAATAAG-----AAGCGTCGC 165  
Db 237 GlnArgArgGlyPheLeuCysAsnGlnAlaProHisLysHisGlyPheLeuGluGlyArg 256  
QY 166 CAGGTGAATTGCTCTTCCTCCGCTACATCTCGGACTGGGACCTAGACCCTGGCGTGC 225  
Db 257 HisAlaGluLeuCysPheLeuAspValIleProPheTrpLysLeuAspGlnAsp 276  
QY 226 TACCGCTCACCTGGTTTCACCTCCTCGGAGCCCTGCTACGACTGTGCCCATCTGGGCC 285  
Db 277 TyrArgValThrCysPheThrSerTrpSerProcysPheSerCysAlaGlnGluMetAla 296  
QY 286 GACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCAACCGCGCGCTCTACTTC 345

Db	297	LysPheIleSerLysasnLysHisValSerLeuCysIlePheThrAlaArgIleYr---	315
QY	346	TGTGAGGACCGCAAGGCTCAGGCCGAGGGCTCGCGGCTCGACCGCGCGGGTGCAA	405
Db	316	---AspAspGlnGlyArgCysGlnGluGlyLeuArgThrLeuAlaGluAlaGlyAlaLys	334
QY	406	ATAGGCATCATGACCTTCAAACTATATTTTACTGCTGGAATACITTTGTAGAAACCAT	465
Db	335	IleSerIleMerThrYrSerCluphelysHisCysTrpAspThrPheValAspHisGln	354
QY	466	GAAGAAGCTTCAAGAGCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCACAGAC	525
Db	355	GlyCysProPheGlnProTrpAspGlyLeuAspGluHisSerGlnAspLeuSerGlyArg	374
QY	526	CTTCGGCGCATCCTT	540
Db	375	LeuArgAlaIleLeu	379

## RESULT 6

RESOLUTIO  
AA84437

ID AAY84437 standard; Protein; 384 AA.

XX  
XX

AC AAY84437;  
yy

DT 25-JUL-2000 (first entry)

Amino acid sequence of a human RNA-associated protein.

XX  
KW  
Human: DNA-20

KW immune response: nonproducing;  
KW Human; RNA-associated protein; cell

[illegible]

WO200015799-A2.

23-MAR-2000.

17-SEP-1999; 99WO-US21688

X  
R 17-SEP-1998; 98US-0156039.





CC This sequence represents the human RNA editing enzyme REE-2. cDNA  
 CC encoding REE-2 was initially isolated in a prostate tumour cDNA library.  
 CC with the cDNA encoding the present sequence representing a consensus.  
 CC REE-2 has chemical and structural homology with the human apoB mRNA  
 CC editing protein HEPR (28% identity), the rat HEPR homologue REPR (30%  
 CC identity), and a portion of the mRNA editing enzyme phosphorin I (43%  
 CC identity). REE-2 was found to be expressed in a variety of cDNA  
 CC libraries, a high proportion of which were derived from tumours, neuronal  
 CC tissues, immune system cells or synovial tissue from arthritis patients.  
 CC REE-2 is therefore thought to be associated with the development of  
 CC cancer, autoimmune disorders, circulatory system disorders (e.g.  
 CC hypercholesterolaemia), viral infections and neurological diseases (e.g.,  
 CC neurofibromatosis). REE-2 or its nucleic acids may be used in the  
 CC diagnosis, treatment and prevention of such diseases via the modulation  
 CC of transcript editing, which in turn has effects on the encoded protein  
 CC (e.g., an alteration in protein activity). The invention specifically  
 CC relates to methods of detecting nucleic acids encoding human REE-2 in a  
 CC biological sample.  
 XX  
 SQ Sequence 190 AA;

Alignment Scores:  
 Pred. No.: 3,72e-33 Length: 190  
 Score: 369.50 Matches: 79  
 Percent Similarity: 58.52% Conservative: 24  
 Best Local Similarity: 44.89% Mismatches: 64  
 Query Match: 33.56% Indels: 9  
 DB: 21 Gaps: 4

US-09-966-880A-7\_COPY\_80\_676 (1-597) x AAB11973 (1-190)

QY 31 TTTCTTTTACCAATTCAAAATGTCGCGTGGCTAGGCTGCGGTGAGACCTACTGTGC 90  
 DB 17 PheTyrPheGlnPheLeuSerProGluAlaAsnAspArgAsnGluThrLeuPheCys 36  
 QY 91 TACGTAGTGAAG-----AGCGGTGACAGTCTACATCTTTTTCACGTGGACTTTGGTAT 144  
 DB 37 PheThrValGluGlyIleLysArgArgSerValSerThrLysThr-----GlyVal 54  
 QY 145 CTTCGCAAT-----AAGACGGCTGCCACGTGGGAATTCCTTCTCCGCTAC 192  
 DB 55 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74  
 QY 193 ATCTCGGACTGGGACCTGACACCTGCGCGTGTACCGCGTACCTGCTTCTCCGCTCG 252  
 DB 75 PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTyrThrSerTrp 94  
 QY 253 AGCCCTGCTACGACTGTGCGCGACATGTGGCGGACTTTCTGCGAGGGAACCCCAAGCTC 312  
 DB 95 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 114  
 QY 313 AGCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAGGCTGAGCCGAG 372  
 DB 115 AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu 133  
 QY 373 GGCTGCGCGGCTGACCGCGCGGTGCAATAGCCATCATGACCTTCAAGATTAT 432  
 DB 134 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrGluAspPhe 153  
 QY 433 TTTTACTGCTGGAATCTTTGTAGAAACCATCAAGAACTTCAAGCCTGGGAGGG 492  
 DB 154 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTyrLysGly 173  
 QY 493 CTGCATGAAATTCAGTCTCTCTCCAGACAGCTTTCGGCGCATCCT 540  
 DB 174 LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu 189

RESULT 9  
 AAM38905  
 ID AAM38905 standard; Protein: 190 AA.  
 XX  
 AC AAM38905;  
 XX

DT 22-OCT-2001 (first entry)  
 XX Human polypeptide SEQ ID NO 2050.  
 DE  
 XX  
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200153312-A1.  
 XX  
 PD 26-JUL-2001.  
 XX  
 PF 26-DEC-2000; 2000WO-US34263.  
 XX  
 PR 21-JAN-2000; 2000US-0498725.  
 PR 25-APR-2000; 2000US-052317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPT; 2001-442253/47.  
 XX  
 DR N-PSDB; AAI38061.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 XX such as central nervous system injuries -  
 XX  
 PS Example 3; SEQ ID NO 2050; 10078pp; English.  
 XX

CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

SQ Sequence 190 AA;

Alignment Scores:

Pred. No.: 3,72e-33 Length: 190  
 Score: 369.50 Matches: 79  
 Percent Similarity: 58.52% Conservative: 24  
 Best Local Similarity: 44.89% Mismatches: 64  
 Query Match: 33.56% Indels: 9  
 DB: 22 Gaps: 4

US-09-966-880A-7\_COPY\_80\_676 (1-597) x AAM38905 (1-190)

QY 31 TTTCTTTTACCAATTCAAAATGTCGCGTGGCTAGGCTGCGGTGAGACCTACTGTGC 90  
 DB 17 PheTyrPheGlnPheLeuSerProGluAlaAsnAspArgAsnGluThrLeuPheCys 36  
 QY 91 TACGTAGTGAAG-----AGCGGTGACAGTCTACATCTTTTTCACGTGGACTTTGGTAT 144  
 DB 37 PheThrValGluGlyIleLysArgArgSerValSerThrLysThr-----GlyVal 54  
 QY 145 CTTCGCAAT-----AAGACGGCTGCCACGTGGGAATTCCTTCTCCGCTAC 192  
 DB 55 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74  
 QY 193 ATCTCGGACTGGGACCTGACACCTGCGCGTGTACCGCGTACCTGCTTCTCCGCTCG 252  
 DB 75 PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTyrThrSerTrp 94  
 QY 253 AGCCCTGCTACGACTGTGCGCGACATGTGGCGGACTTTCTGCGAGGGAACCCCAAGCTC 312  
 DB 95 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 114  
 QY 313 AGCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGACCGCAGGCTGAGCCGAG 372  
 DB 115 AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu 133  
 QY 373 GGCTGCGCGGCTGACCGCGCGGTGCAATAGCCATCATGACCTTCAAGATTAT 432  
 DB 134 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrGluAspPhe 153  
 QY 433 TTTTACTGCTGGAATCTTTGTAGAAACCATCAAGAACTTCAAGCCTGGGAGGG 492  
 DB 154 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTyrLysGly 173  
 QY 493 CTGCATGAAATTCAGTCTCTCTCCAGACAGCTTTCGGCGCATCCT 540  
 DB 174 LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu 189



Db	17	PheTyrPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys	36	PR	11-JUL-2000;	2000US-0217496.
QY	91	TACGTAGTGAAG-----AGGCGTGACAGTGCTACATCTTTCACTGGACTTGGTTAT	144	PR	14-JUL-2000;	2000US-0218290.
Db	37	PheThrValGluGlyIleLysArgSerValValSerTrpLysThr-----GlyVal	54	PR	26-JUL-2000;	2000US-0220963.
QY	145	CTTCGCAAT-----AAGAACGGCTGCCACGTGGAATTGCTTCTCCCTCCGCTAC	192	PR	26-JUL-2000;	2000US-0220964.
Db	55	PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp	74	PR	14-AUG-2000;	2000US-0224518.
QY	193	ATCTCGGACTGGACCTAGACCTGCGCGCTGCTACCGGTGCTACCTGGTTCACCTCCTGG	252	PR	14-AUG-2000;	2000US-0224519.
Db	75	PheCysAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTyrThrSerTrp	94	PR	14-AUG-2000;	2000US-0225213.
QY	253	AGCCCTGCTACGACTGTGCCGACATGTGGCGGACTTCTCGGAGGAACCCCAACCTC	312	PR	14-AUG-2000;	2000US-0225214.
Db	95	SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal	114	PR	14-AUG-2000;	2000US-0225267.
QY	313	AGTCTGAGGATCTACCGCGCCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAG	372	PR	14-AUG-2000;	2000US-0225268.
Db	115	AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu	133	PR	14-AUG-2000;	2000US-0225270.
QY	373	GGGCTCGGGGTGCACCGCGCGGGTGCAATAGCCATCATGACCTTCAAGATTAT	432	PR	14-AUG-2000;	2000US-0225271.
Db	134	GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrGluAspPhe	153	PR	14-AUG-2000;	2000US-0225275.
QY	433	TTTTACTGCTGGAATACTTTGTAGAAACCATGAAAGACTTTCAAAGCCTGGGAAGG	492	PR	14-AUG-2000;	2000US-0225278.
Db	154	LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTyrLysGly	173	PR	18-AUG-2000;	2000US-0225759.
QY	493	CTGCATGAAATTCAGTCTCGTCTCTCCAGACACTTCGGCGCATCCTT	540	PR	22-AUG-2000;	2000US-0226681.
Db	174	LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu	189	PR	22-AUG-2000;	2000US-0226686.
RESULT 10				PR	22-AUG-2000;	2000US-0227182.
AAU23350				PR	30-AUG-2000;	2000US-0228924.
ID	AAU23350	standard; Protein; 210 AA.		PR	01-SEP-2000;	2000US-0229287.
AC	AAU23350;			PR	01-SEP-2000;	2000US-0229343.
XX				PR	01-SEP-2000;	2000US-0229344.
DT	17-DEC-2001	(first entry)		PR	01-SEP-2000;	2000US-0229345.
DE				PR	05-SEP-2000;	2000US-0229509.
XX				PR	05-SEP-2000;	2000US-0229513.
KW	Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;			PR	06-SEP-2000;	2000US-0230437.
KW	ligase; hyperproliferative disorder; immunodeficiency disorder;			PR	08-SEP-2000;	2000US-0230438.
KW	autoimmune disorder; neurological disorder; metabolic disorder;			PR	08-SEP-2000;	2000US-0231242.
KW	inflammatory disorder; cardiovascular disorder; reproductive disorder;			PR	08-SEP-2000;	2000US-0231243.
KW	blood-related disorder; infectious disorder; cytostatic; anti arthritic;			PR	08-SEP-2000;	2000US-0231244.
KW	nephrotropic; anticoagulant.			PR	08-SEP-2000;	2000US-0231413.
OS	Homo sapiens.			PR	08-SEP-2000;	2000US-0231414.
XX				PR	08-SEP-2000;	2000US-0232080.
PN	WO200153301-A2.			PR	08-SEP-2000;	2000US-0232081.
XX				PR	12-SEP-2000;	2000US-0231968.
PD	02-AUG-2001.			PR	14-SEP-2000;	2000US-0232397.
XX				PR	14-SEP-2000;	2000US-0232398.
PF	17-JAN-2001;	2001WO-US01239.		PR	14-SEP-2000;	2000US-0232399.
XX				PR	14-SEP-2000;	2000US-0232400.
PR	31-JAN-2000;	2000US-0179065.		PR	14-SEP-2000;	2000US-0232401.
PR	04-FEB-2000;	2000US-0180628.		PR	14-SEP-2000;	2000US-0233063.
PR	24-FEB-2000;	2000US-0184664.		PR	14-SEP-2000;	2000US-0233064.
PR	02-MAR-2000;	2000US-0186350.		PR	14-SEP-2000;	2000US-0233065.
PR	16-MAR-2000;	2000US-0189874.		PR	21-SEP-2000;	2000US-0234223.
PR	17-MAR-2000;	2000US-0190076.		PR	21-SEP-2000;	2000US-0234274.
PR	18-APR-2000;	2000US-0198123.		PR	25-SEP-2000;	2000US-0234997.
PR	19-MAY-2000;	2000US-0205515.		PR	25-SEP-2000;	2000US-0234998.
PR	07-JUN-2000;	2000US-0209467.		PR	26-SEP-2000;	2000US-0235484.
PR	28-JUN-2000;	2000US-0214886.		PR	27-SEP-2000;	2000US-0235834.
PR	30-JUN-2000;	2000US-0215135.		PR	27-SEP-2000;	2000US-0235836.
PR	07-JUL-2000;	2000US-0216647.		PR	29-SEP-2000;	2000US-0236327.
PR	07-JUL-2000;	2000US-0216880.		PR	29-SEP-2000;	2000US-0236367.
PR	11-JUL-2000;	2000US-0217487.		PR	29-SEP-2000;	2000US-0236369.
				PR	29-SEP-2000;	2000US-0236370.
				PR	02-OCT-2000;	2000US-0236802.
				PR	02-OCT-2000;	2000US-0237037.
				PR	02-OCT-2000;	2000US-0237038.
				PR	02-OCT-2000;	2000US-0237039.
				PR	13-OCT-2000;	2000US-0237040.
				PR	13-OCT-2000;	2000US-0239935.
				PR	13-OCT-2000;	2000US-0239937.
				PR	20-OCT-2000;	2000US-0240960.
				PR	20-OCT-2000;	2000US-0241221.
				PR	20-OCT-2000;	2000US-0241785.
				PR	20-OCT-2000;	2000US-0241786.
				PR	20-OCT-2000;	2000US-0241787.
				PR	20-OCT-2000;	2000US-0241808.
				PR	20-OCT-2000;	2000US-0241809.
				PR	20-OCT-2000;	2000US-0241826.



KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
KW Leukaemia.  
XX  
OS Homo sapiens.  
XX  
XX WO200153312-A1.  
PN  
XX  
XX 26-JUL-2001.  
PD  
XX  
XX 26-DEC-2000; 2000WO-US34263.  
PF  
XX  
XX 21-JAN-2000; 2000US-0488725.  
PR  
XX 25-APR-2000; 2000US-0552317.  
PR  
XX 09-JUL-2000; 2000US-0598042.  
PR  
XX 19-JUL-2000; 2000US-0620312.  
PR  
XX 03-AUG-2000; 2000US-0653450.  
PR  
XX 14-SEP-2000; 2000US-0653450.  
PR  
XX 19-OCT-2000; 2000US-0662191.  
PR  
XX 29-NOV-2000; 2000US-0693036.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Zhao J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
PI  
XX  
DR WPI; 2001-442253/47.  
DR N-PSDB; AAI59847.  
XX  
XX Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -  
PT  
XX  
XX Example 2; SEQ ID NO 5622; 10078pp; English.  
PS  
XX  
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
CC in gene therapy. A composition containing a polypeptide or polynucleotide  
CC of the invention may be used to treat diseases of the peripheral nervous  
CC system, such as peripheral nervous injuries, peripheral neuropathy and  
CC localised neuropathies and central nervous system diseases, such as  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
CC utilisation of the activities such as: Immune system suppression,  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
CC assays for receptor activity, arthritis and inflammation, leukaemias and  
CC C.N.S disorders.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification.  
XX  
XX  
SQ Sequence 221 AA;

Alignment Scores:  
Pred. No.: 3,93e-33 Length: 221  
Score: 369.50 Matches: 79  
Percent Similarity: 58.52% Conservative: 24  
Best Local Similarity: 44.89% Mismatches: 64  
Query Match: 33.56% Indels: 9  
DB: 22 Gaps: 4

US-09-966-880A-7\_COPY\_80\_676 (1-597) x AAM40691 (1-221)

QY 31 TTCTCTTACCAATTCAAATATGTCGGCTAAGGTCGGCGTGAGACCTACCTGTGC 90  
Db 48 PheTyrPheGlnPheLysAsnLeuTyrGluAlaAsnAspArgAsnGluThrTrpLeuCys 67  
QY 91 TAGGTAGTGAAG-----AGGCGTGACAGTGCTACATCCCTTTTCACTGGACTTTGGTTAT 144  
Db 68 PheThrValGluGlylleLysArgSerValValSerTrpLysThr-----GlyVal 85  
QY 145 CTTCCGAAT-----AAGAACGGTGCCAGCTGGAATTCCTTCCCTCCGCTAC 192

Db 86 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 105  
QY 193 ATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGCTCACCCTGCTTACCTCCCTGG 252  
Db 106 PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTrpThrSerTrp 125  
QY 253 AGCCCTCTGCTACGACTGTGCGCGACATGTGCGCGCACTTCTCTCGAGGAGCAACCTC 312  
Db 126 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 145  
QY 313 AGTCTGAGGATCTTACCGCGCGCTCTACTCTGTGTGTGAGGACCGCAAGCTGAGCCCGAG 372  
Db 146 AsnLeuThrIlePheThrAlaArgLeuTyrTrpPheGln---TyrProCysTyrGlnGlu 164  
QY 373 GGGCTCGGGCGCTGCACCGCGCGGTGCAATAGCATCATCATCATCATCATCATCATCAT 432  
Db 165 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrIleAspPhe 184  
QY 433 TTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTTCAAGCCCTGGGAAGG 492  
Db 185 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly 204  
QY 493 CTGATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGGATCCCT 540  
Db 205 LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu 220

RESULT 12  
AAB57061  
ID AAB57061 standard; Protein; 222 AA.  
XX  
AC AAB57061;  
XX  
DT 13-MAR-2001 (first entry)  
XX  
DE Human prostate cancer antigen protein sequence SEQ ID NO:1639.  
XX  
KW Human; prostate cancer; prostate cancer antigen; detection; diagnosis;  
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;  
KW vulnar; gastrotentinal; nephrotropic; antinfetive; gynaecological;  
KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
KW wound; infectious disease.  
XX  
OS Homo sapiens.  
XX  
XX WO200055174-A1.  
PN  
XX  
XX 21-SEP-2000.  
PD  
XX  
XX 08-MAR-2000; 2000WO-US05988.  
XX  
XX 12-MAR-1999; 99US-0124270.  
PR  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA (ROSE/) ROSEN C A.  
XX  
XX Rosen CA, Ruben SM;  
PI  
XX WPI; 2000-587513/55.  
DR  
XX N-PSDB; AAF16264.  
DR  
XX  
XX Prostate cancer associated gene sequences, referred to as prostate  
PT cancer antigens, useful for treatment, prevention, and diagnosis of  
PT disorders such as prostate cancer -  
XX  
XX Claim 11; Page 2097-2098; 2338pp; English.  
XX  
XX AAF15566 to AAF16505 encode the human prostate cancer associated  
CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
CC The prostate cancer antigens can have neuroprotective, cytostatic,  
CC cardioactive, immunomodulatory, muscular, vulnar, gastrotentinal,  
CC nephrotropic, antinfetive, gynaecological and antibacterial activities,  
CC and can be used in gene therapy. The prostate cancer antigen

CC polynucleotides may be used for detection of prostate cancer, chromosome  
CC identification, as chromosome markers, and for numerous other diagnostic  
CC or research purposes. The prostate cancer antigens may be used to treat  
CC disorders such as neural, immune, muscular, reproductive,  
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
CC disorders, wounds, and infectious diseases. AAF16514 to  
CC AAB57303 represent sequences used in the exemplification of the present  
CC invention.

XX  
SQ Sequence 222 AA;  
Alignment Scores:  
Pred. No.: 3,93e-33 Length: 222  
Score: 369.50 Matches: 79  
Percent Similarity: 58.52% Conservative: 24  
Best Local Similarity: 44.89% Mismatches: 64  
Query Match: 33.56% Indels: 9  
DB: 21 Gaps: 4

US-09-966-880A-7\_COPY\_80\_676 (1-597) x AAB57061 (1-222)

QY 31 TTCTTTTACCAATTCAAAATGTCGCTGGGCTAAGGGTGGGCTGAGACCTACCTGTC 90  
DB 49 PhETyrPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuGly 68  
QY 91 TAGCTAGTGAAG-----AGCGCTGACAGTGTACATCCTTTTCACTGGACTTTGGTTAT 144  
DB 69 PheThrValGluGlyLeuLysArgArgSerValValSerTrpLysThr-----GlyVal 86  
QY 145 CTTTCGCAT-----AAGACGGCTGCCAGTGAATGCTTCTTCCTCCGCTAC 192  
DB 87 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 106  
QY 193 ATCTCGGACGTGGGACCTAGACCTGGCGCTGCTACCGCTACCTGGTTCACCTCCTCG 252  
DB 107 PheCysAspAspLeuSerProAsnThrLysTyrGlnValThrTrpTrpThrSertTrp 126  
QY 253 AGCCCTGTCTACGACTGTGCCGACATGTGGCGGACTTTCTGGCGGGAACCCCAACCTC 312  
DB 127 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 146  
QY 313 AGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAG 372  
DB 147 AsnLeuThrLlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu 165  
QY 373 GGCTGGCGGCTGTGACCGCGCGGCTGCTCAATAGCCATCATGACCTTCAAGATTAT 432  
DB 166 GlyLeuArgSerLeuSerGlnGlyValAlaValAlaGluLeuMetAspTyrGluAspPhe 185  
QY 433 TTTTACTGCTGGAATCTTTTGTAGAAACCATCAAGAACTTTCAAGCCTGGGAGGG 492  
DB 186 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly 205  
QY 493 CTGCATGAAATTCAGTTCTCTCCAGACACCTTCGGGGCATCCTT 540  
DB 206 LeuLysThrAsnPheArgLeuLeuLysArgArgLeuGluSerLeu 221

RESULT 13

AAU23537  
ID AAU23537 standard; Protein: 222 AA.

XX AC AAU23537;

XX DT 17-DEC-2001 (first entry)

XX DE Novel human enzyme polypeptide #623.

XX KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KW KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;  
KW KW nephrotropic; anticoagulant.

XX  
OS Homo sapiens.  
XX WO200155301-A2.  
XX  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01239.  
XX  
XX PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226688.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 01-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.





XX OS Homo sapiens.  
XX PN WO200155301-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01239.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX

PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX WPI: 2001-465566/50.

DR N-PSDB; AAS41669.

XX Novel polypeptides and polynucleotides useful for diagnosing,

PT preventing, treating neural, immune system, muscular, reproductive,

PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous

XX diseases

PS Claim 11; SEQ ID No 1795; 1180pp; English.

XX The present invention relates to the isolation of novel human enzyme  
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences  
CC encoding them. The enzyme polypeptides of the invention may comprise the  
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
CC isomerases or ligases. The sequences of the invention are useful in the  
CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
CC disorders including hyperproliferative disorders (e.g. cancer),  
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders  
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
CC blood-related disorders (e.g. haemophilia), reproductive disorders  
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The  
CC polynucleotides of the invention can also be used in gene therapy.  
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the  
CC invention.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 272 AA;

Alignment Scores:

Pred. No.:	7.39e-28	Length:	272
Score:	323.50	Matches:	75
Percent Similarity:	55.25%	Conservative:	25
Best Local Similarity:	41.44%	Mismatches:	67
Query Match:	29.38%	Indels:	14
DB:	22	Gaps:	6

US-09-966-880A-7\_COPY\_80\_676 (1-597) x AAU23799 (1-272)

QY	31	TTTTCTTTACAAATGTCGCTGGGCTAAGGTCGGCGTGAGACCTACCTGTGC	90
DB	49	PhetYrPheGlnPhelysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys	68
QY	91	TACGTAGTGAAG-----AGGGCTGACAGTGCTACATCCTTTTCTACCTGGACTTTGGTTAT	144
DB	69	PhetHrValGluGlyIleLysArgSerValValSerTrpLysThr-----GlyVal	86
QY	145	CTTCGCAAT-----AAGAACGGCTGCCAGCTGGAAATGCTCTTCCTCCGCTAC	192
DB	87	PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp	106
QY	193	ATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGCTGCACCTGGTTACCTCCCTCG	252
DB	107	PheCysAspAspIleLeuSerProAsnThrLysTyrglnValThrTrpTrpThrSerTrp	126
QY	253	AGCCCTGTACGACTGTCCCGACATGTGGCCGACTTTCTCGAGGGAGACCCCAACCTC	312
DB	127	SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal	146
QY	313	AGTCTGAGGATCTTACCCGGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAG	372
DB	147	AsnLeuThrIlePheHrAlaArgLeuTyrtyrPheGln---TyrProCysTyrglnGlu	165
QY	373	GGGCTCGCGGGCTGACCGCGCGGTGCAATAGCCATCATGACCTTCAAGATTAT	432
DB	166	GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrgluAspPhe	185

\* . \*

QY	433	TTTTACTGCTGGAATACTTTTGTAGAAAAACCATGAAAGAACTTTCAAAGCCTGGGAAGG	492
DB	186	LysTyrCysTrpGluAsnPheValTyrAsnAsp-AsnGlu-----ProAlaProly	202
QY	493	CTGCATGAAAATTCAGTTTCGTC---TCTCAGACACACTTTCGGCGCATCTTTTGGCCCTG	549
DB	202	scysLeuSerLeuValLeuThrArgGlyCysSerSerGlnAlaAlaGlyTrpLeuProcy	222
QY	550	T 550	
DB	222	s 222	

Search completed: June 14, 2003, 18:25:30  
Job time : 28.5713 secs



GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 14, 2003, 18:12:33 ; Search time 17.1832 Seconds

Title: US-09-966-880A-7\_COPY\_80\_676  
Perfect score: 1101  
Sequence: 1 atggacacgctcttgatgaa.....ttcgtactttggactttga 597  
6680.033 Million cell updates/sec

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-O=/cpn2.1/USPTO\_spool/US09966880/runat\_14062003\_175525\_10329/app\_query.fasta\_1.9493  
-DB=PIR\_73 -QFWT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09966880.@CGN\_1.1.456.@runat\_14062003\_175525\_10329 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMECUT=120 -WARN\_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PIR\_73:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description	
1	245	22.3	116	2	G01233	phorbolin I - huma	
2	225	20.4	236	2	A53853	apolipoprotein B m	
3	216.5	19.7	236	2	I59323	apolipoprotein B m	
4	211	19.2	229	2	JC4269	apolipoprotein B m	
5	211	19.2	229	2	I48249	apolipoprotein B m	
6	198	18.0	229	2	I59577	apolipoprotein B m	
c	7	97.5	9.4	295	2	S58850	homeotic protein a
8	97.5	8.9	369	2	S63464	hypothetical prote	
9	94.5	8.6	568	2	JC7317	cyclin-dependent k	
c	10	94	9.1	504	2	JC1306	virion protein hom
11	93	8.4	277	2	A46241	interferon respons	
12	92.5	8.4	2342	2	T13412	hypothetical prote	
13	91.5	8.3	839	2	T04859	extensin homolog F	
14	88.5	8.0	1493	2	A38218	GAP-associated pro	

ALIGNMENTS

RESULT 1  
G01233  
Phorbol I - human (fragment)  
C;Species: Homo sapiens (man)  
C;Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 19-Apr-2002  
C;Accession: G01233  
R;Madson, P.P.  
submitted to the EMBL Data Library, December 1993  
A;Reference number: G06330  
A;Accession: G01233  
A;Status: preliminary; translated from GB/EMBL/DBD  
A;Molecule type: mRNA  
A;Residues: 1-116 <MAD>  
A;Cross-references: EMBL:U03891; NID:g436940; PIDN:AAA03706.1; PID:g436941  
C;Superfamily: apolipoprotein B mRNA editing enzyme, catalytic chain 1

Alignment Scores:  
Pred. No.: 2.31e-15 Length: 116  
Score: 245.00 Matches: 51  
Percent Similarity: 61.82% Conservative: 17  
Best Local Similarity: 46.36% Mismatches: 32  
Query Match: 22.25% Indels: 10  
DB: 3 Gaps: 3

US-09-966-880A-7\_COPY\_80\_676 (1-597) x G01233 (1-116)

QY	226	TACCGCGTCACCGTTCACCTCTCGAGCCCTGCTACGAC-----TGTGCCCGACAT	279
Db	7	TyrArgValThrTrpPheIleSerTrpSerProcysPheSerTrpGlyCysAlaGlyGlu	26
QY	280	GTGCGCGACVTTCTGCGAGGAAACCCAACTCAGTCTGAGGATCTTCACCGCGCGCTC	339
Db	27	ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgIle	46
QY	340	TACTTCTGTGAGGACCGCAAGCTGAGCCC-----CAGGGGCTGGCGGCGCTGCAC	390
Db	47	Tyr-----AspTyrAspProLeuTyrIysGluAlaLeuGlnMetLeuArg	61

391	CGCCGCCGGGTGCAATAGCCATCATGACCTTCAAGATTATTTTACTCGTGAATACT	450
QY		
Db		
62	ASPAAGAGAGAGValSerIleMetThrTyrAspGluPheGluTyrCysTrpAspThr	81
451	TTTTGTAGAAACCATGAAGAACTTTCAAACGCTGGGAAGGGCTGCATGAAATTCAGTT	510
QY		
82	PheValTyrArgGlnGlnGlyCysProPheGlnProTyrAspGlyLeuGluGluHisSerGln	101
Db		

## RESULT 2

apoliipoprotein B mRNA editing enzyme, catalytic chain 1 (EC 3.5.4.-) - rabbit C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:date: 07-Oct-1994 #sequence\_revision 07-Oct-1994 #text\_change 18-Aug-2000  
C:Accession: A53853  
R:Yamanaka, S.; Poksay, K.S.; Balestra, M.E.; Zeng, G.Q.; Innerarity, T.L.  
J. Biol. Chem. 269, 21725-21734, 1994  
A:Title: Cloning and mutagenesis of the rabbit ApoB mRNA editing protein. A Z

Alignment Scores:			
Pred. No.:	1,99e-13	Length:	236
Score:	225.00	Matches:	49
Percent Similarity:	57.2%	Conservative:	26
Best Local Similarity:	37.4%	Mismatches:	48
Query Match:	20.4%	Indels:	8
DB:	2	Gaps:	4

US-09-966-880A-7\_COPY\_80\_676 (1-597) x A53853 (1-236)

QY	70	CGCGCTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTGTACATCTCTTTTCA	129
Db	33	ArglyysGluAlaCysLeuLeuYrGluLeuYrGlyAlaSerSerlySthrTrpArg	52
QY	130	CTGGACTTGGTTATCTCTCCGAATAAGAACGGGTGC---CACGTGGAATGTCTTCCTC	186
Db	53	SerSer-----GlyLyAsnThrThrAsnHisValGluValAsnPheLeu	67
QY	187	---CCGTACATCTCGGACTGGGACCTAGACCTGGCGGTGTACCGCGTGCACCTGGTTC	243
Db	68	GlulysLeuThrSerGluGlyArgLeuGlyProSerThrCysSerIleThrTrpPhe	87
QY	244	ACCTCTGGAGCCCTCTACGACTGTGCCCGACATGTGCCGACATTTCTGGCAGGGAAC	303
Db	88	LeuSerTrpSerProCysTrpGluCysSerMetAlaIleArgGluPheLeuSerGlnHis	107
QY	304	CCCAACCTCAGCTGAGGATCTTACCGCGCGGCTCTACTTCTGTGAGACCGCAAGGCT	363
Db	108	ProGlyValThrLeuIlePheValAlaArgLeuPheGlnHisMetAspArgArg---	126
QY	364	GAGCCGAGGGCTGCGCGGCTGCACCGCGGGGTGCAATACCATGCCATGACCTTC	423
Db	127	AsnArgGlnGlyLeuLysAspLeuValThrSerGlyValThrValArgValMetSerVal	146
QY	424	AAAGATTATTTTACTGCTCGAATACTTTGTA	456
Db	147	SerGluYrCysYrCysTrpGluAsnPheVal	157

### RESULT 3

apolipoprotein B mRNA editing enzyme, catalytic chain 1 (EC 3.5.4.-) - human  
C:Species: Homo sapiens (man)

C:Date: 31-May-1996 #sequence\_revision 31-May-1996 #text\_change 18-Aug-2000  
C:Accession: I59323; S45253  
R:Lau, P.; Zhu, H.J.; Baldini, A.; Charnsangavej, C.; Chan, L.  
Proc. Natl. Acad. Sci. U.S.A. 91, 8522-8526, 1994  
A:title: Dimeric structure of a human apolipoprotein B mRNA editing protein  
A:Reference number: I59323; MUID: h4359963; PMID:8078915  
A:Accession: I59323  
A>Status: translated from GB/EMBL/DBJ

Alignment Scores:	1.3e-12	Length:	236
Pred. No.:	216.50	Matches:	44
Score:	59.53%	Conservative:	31
Percent Similarity:	34.93%	Mismatches:	42
Best Local Similarity:	19.66%	Indels:	9
Query Match:	2	Gaps:	4
DB:	2		

US-09-966-880A-7\_COPY\_80\_676 (1-597) x I59323 (1-236)

[illegible]

## RESULT 4

apolipoprotein B mRNA editing enzyme, catalytic chain 1 (EC 3.5.4.-) - mouse  
J04289

C;Accession: J04269  
R;Osuga, J.; Inaba, T.; Harada, K.; Yagyu, H.; Shimada, M.; Yazaki, Y.; Yamada, N.; Ishii, K. (1995) Biochem. Biophys. Res. Commun. 214, 653-662, 1995  
A;Title: Cloning and structural analysis of the mouse apolipoprotein B mRNA editing protein  
A;Reference number: J04269; M01D:95408299; PMID:7677778

A;Molecule type: mRNA  
A;Residues: 1-229 <OSU>  
C;Comment: This protein belongs to the cytidine deaminase gene family.  
C;Superfamily: apolipoprotein B mRNA editing enzyme, catalytic chain 1  
C;Keywords: hydrolase; zinc finger  
F;48-96/Region: zinc finger HHCC motif

Alignment Scores:	
Pred. No.:	4, 4e-12
Score:	211.00
Percent Similarity:	55.30%
Best Local Similarity:	37.12%
Query Match:	19.16%
DB:	2
Length:	229
Matches:	49
Conservative:	24
Mismatches:	49
Indels:	10
Gaps:	4

US-09-966-880A-7\_COPY\_80\_676 (1-597) x JC4269 (1-229)

Qy 70 CGCGGTGAGACCTACTCTGTGCTACGTAGTGAAG-----AGGCGTGACAGTGCCTACATCC 123  
||||:||||| ||| ||| ::||| ||| |||

Db 33 ArgLysGluThrCysLeuLeuTyrGluIleAsnTrpGlyGlyArgHisSerVal----- 50

Qy 124 TTTTCACGTGGACTTTTGGTTATCTTCCACATTAAGAACGGCTGCCACGTGGGAATTGCTCTTC 183

```

00 -----iipargh1stnrserglnasnrserasnh1svalguvalasphne 66
31 -----
104 nmc -----cccccccccccccccccccccccccccccccccccccccccccc
000 -----cccccccccccccccccccccccccccccccccccccccccccccccccc

```

104 C C C T A C A T C T C G G A C T G G G A C C T A G A C C C T G G C G G T C A C C T G G 240

6/ LeuGluLysPheThrTrnGluArgTyrPheArgProAsnTrnArgCysSerIleThrTrp 86

241 TTTCACCTCCGGAGCCCCGCTACGACTGTGCCGCACATGTGCCGACTTTCGCGAGG 300

D6 87 PheLeuSerTrpSerProCysGlyGluCysSerArgAlaIleThrGluPheLeuSerArg 106

QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360

Db 107 HisProTyrValThrLeuPheIleTyrIleAlaArgLeuTyrHisHisThrAspGlnArg 126

Qy 361 GCTGACCCGAGGGGTGCGGGGCTGCACCGCGGGGGTGCAAATAGCCATCATGACC 420

Db 127 ---AsnArgGlnGlyLeuArgAspLeuIleSerSerGlyValThrIleGlnIleMetThr 145

Qy 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTGTA 456

Db 146 GluGlnGluTyrCysTyrCysTrpArgAsnPheVal 157

RESULT 5  
T482A9

apolipoprotein B mRNA editing enzyme, catalytic chain 1 (EC 3.5.4.-) - mouse

```
C;species: mus musculus (house mouse)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Aug-2000
```

R; Nakamura, M.; Oka, K.; Krushkal, J.; Kobayashi, K.; Yamamoto, M.; Li, W.H.; Chan, L.  
C; Accession: I48249; I49287; I49286; I49288

**A:**Title: Alternative mRNA splicing and differential promoter utilization determine tissue

ion of Apobec1 and related nucleoside/nucleotide deaminases.  
A; Reference number: A57020; MUID:95286585; PMID:7768898

A;Accession: I48249  
A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA  
A;Residues: 1-229 <RES>

A;Cross-references: EMBL:U21951; NID:g899501; PIDN:AAC52211.1; PID:g899503  
A;Accession: I49287

A; Status: preliminary; translated from GB/EMBL/DBBJ  
A; Molecule type: mRNA

A; Residues: 1-229 <RE2>  
A: Cross-references: EMBL:U22263. NID:q890506. PIDN:AAC52313 1. PID:q800507

**A;Accession:** I49286

Pred. No.:	7.82e-11	Length:	229
Score:	198.00	Matches:	56
Percent Similarity:	47.0%	Conservative:	24
Best Local Similarity:	32.94%	Mismatches:	62
Query Match:	17.98%	Indels:	28
DB:	2	Gaps:	5

US-09-966-880A-7\_COPY\_80\_676 (1-597) x I59577 (1-229)

QY	70	CGCGTGAGACCTACCTGTGCTACGTAGTAGAAG-----AGG	105
Db	33	ArgLysGluThrCysLeuLeuTyrGluIleAsnTrpGlyClyArgHisSerIleTrpArg	52
QY	106	CGTGACAGTGCTACATACCTTTTTCACGTGGACTTTGGTTATCTTCGCGCAATGAAGACGGCTGC	165
Db	53	HisThrSerGlnAsnThr-----AsnLys-----	60
QY	166	CACGTGGAATTGCTCTTCTCTC-----CGCTACATCTCGGACTGGACCTAGACACCTGGCCGC	222
Db	61	HisValGluValAsnPheIleGluLysPheThrThrGluArgTyrPheCysProAsnThr	80
QY	223	TGCTACCGCGTGACCTGGTTCACCTCCTCGAGGCCCTGCTACGACTGTGCCGACATGTG	282
Db	81	ArgCysSerIleThrTrpPheLeuSerTrpSerProCysGlyGluCysSerArgAlaIle	100
QY	283	GCGGACTTTCGCGAGGAACCCAAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTAC	342
Db	101	ThrGluPheLeuSerArgTyrProHisValThrLeuPheIleTyrIleAlaArgLeuTyr	120
QY	343	TTCTGTGAGGACCGAAGGCTGAGCCGAGGGGCTGCGGGCTGCACCGCGCGGGGTG	402
Db	121	HisHisAlaSerProArg---AsnArgGlnGlyLeuArgAspLeuIleSerSerGlyVal	139
QY	403	CAAAATGACCATGACACCTTCACGAATTATTTTACTGCTGCGAATCTTTTGTAGAAAC	462
Db	140	ThrIleGlnIleMetThrGluGlnGluSerGlyTyrCysTrpArgAsnPheValAsnTyr	159
QY	463	CATGAAGAAGACTTTTCAACCGCTGGGAGGGCTGCATGAAATTCAGTTGCTCTCCAGA	522
Db	160	SerProSerAsnGluAlaHisTrpProArgTyrProHisLeuTrpValArgLeu-----	177
QY	523	CACCTTCGCGGATCCTTTTGGCCCTGTAT	552
Db	178	-----TyrValLeuGluLeuTyr	183

RESULT 7  
homeotic protein abd-A - Junonia coenia  
S58850  
N:Alternate names: abdominal-A homeodomain protein  
C:Species: Junonia coenia  
C:Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 17-Nov-2000  
C:Accession: S58850; S58851  
R:Warren, R.W.; Nagy, L.; Selegue, J.; Gates, J.; Carroll, S.  
Nature 372, 458-461, 1994  
A:Title: Evolution of homeotic gene regulation and function in flies and butterflies.  
A:Reference number: S58850; MUID:95075456; PMID:7840822  
A:Accession: S58850  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-295 <W&R>  
A:Cross-references: EMBL:L41931; NID:G797276; PIDN:AAA68460.1; PID:G797277  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995  
A:Accession: S58851  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 214-271 <W&Z>  
A:Cross-references: EMBL:L42135; NID:G833750; PIDN:AAA68461.1; PID:G833751  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1995  
C:Genetics:  
A:Gene: abd-a  
C:Superfamily: unassigned homeobox proteins; homeobox homology  
C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:215-271/Domain: homeobox homology <HOX>



```
Best Local Similarity: 27.50% Mismatches: 41
Query Match: 9.09% Indels: 65
DB: 2 Gaps: 7

US-09-966-880A-7_COPY_80_676 (1-597) x JC1306 (1-504)

QY 569 CGTAAGTCATCAACCTCATACAGGGGCAAA-----AGGATGCGC 531
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 361 ArgSerGlySerThrArgThrArgGlyArgAlaAlaArgSerThrThrArgLeuGln 380
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 530 CGAAGCTGCTGGAGAGAGAACTGAATTTTCATGC-----AGCCCTTCCCAAGGCT 480
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 381 ArgProCysGlyProArgArgAlaAlaLysCysArgAlaThrProArgGlnArg 400
   ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 479 TTGAAGTTCTTTTCATGGTTTTCTACAAAAGTATTCACGACGTAAATAATCTTTGAAG 420
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 401 LeuArg----- 402

QY 419 GTCATGATGGCTATTGTCACCCCGCGCGGTGCGAGCCGCGGCTCAGCC 360
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 403 -----AlaArgGlyGluProArgHisThrSerGlySerGly 414

QY 359 TTGCGGTCTCACAGAAGTAGAGCGCGGTGAAGATCCTCAGACTGAGGTGGGTTTC 300
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 415 AlaPheSerGlnGly-----ArgArgProGlyArgValCysArgLeu-GlyTrpAlaCy 432

QY 299 CCTCGCAGAAAGTCGGCCA-----CATGT 276
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 432 slysAlaArgSerGlyProAlaArgGlyGlyProGlyProSerProValArgSerGlyLe 452

QY 275 CGGGCACATGCTAGCAGGGGCTCCAGAGAGGTGAACCGAGTGCAGCGGTAGCAGCGCCA 216
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 452 uGlyLeuSerArgAlaArgGlySerPro----- 461

QY 215 GGCTCTAGTCCAGTCCGAGTGTAGCGGAGGAGAGCAATTCACGTGGCAGCGGT 158
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 462 -----GlyProGlyProAlaCys---GlyGlyProSerArgAlaArgGlyGlyArg 477

RESULT 11
A46241
Interferon response element-binding factor IREBF-2 - mouse (fragment)
C:Species: Mus musculus (house mouse)
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 05-Nov-1999
C:Accession: A46241
R:Yan, C.; Tamm, I.
Proc. Natl. Acad. Sci. U.S.A. 89, 8859-8863, 1992
A:Title: Molecular cloning and characterization of additional factors that bind to the
A:Reference number: A46241, MUID:93028366; PMID:1409578
A:Accession: A46241
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-277 <YAN>
A:Cross-references: GB:L13610; NID:g293674; PIDN:AAA39332.1; PID:g293675
A>Note: sequence extracted from NCBI backbone (NCBIN:115299, NCBIF:115300)

Alignment Scores:
Pred. No.: 0.98 Length: 277
Score: 93.00 Matches: 42
Percent Similarity: 33.11% Conservative: 7
Best Local Similarity: 28.38% Mismatches: 46
Query Match: 8.45% Indels: 53
DB: 2 Gaps: 6

US-09-966-880A-7_COPY_80_676 (1-597) x A46241 (1-277)

QY 131 TGGACTTTGGTTATCTTCCATATAAGACGGCTGCCACGTGGAAATGCTCTCTCCCGCT 190
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 118 TrpThrProTrpThrGlyArgCysSerThrAlaAlaSerTyrglyCysArgTrpArgAla 137
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 191 ACATCTCGGACTGGACCTAGACCTGGCGCTGTACCGCGTCACT----- 238
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138 ThrAlaAlaArgProThrArgThrAlaAlaGlyAlaArgHisProValGlyThrAla 157
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 238 ----- 238
Db 158 AlaAlaAlaThrGlyGlyAlaAlaLeuGlyAspAlaGlyAlaAlaAspLeuGlu 177
QY 239 ---GGTTACCTCCCTCGA-----GCCCTGTACGACTGTGCCCCACATGTGGCGC 286
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 178 AlaGlyAlaGlyProGlyArgGluAlaAlaProAlaAlaThrAlaAlaProSerLeuGlyPro 197
QY 287 ACTTTCTGCGGAGGAACCC---CAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACT 343
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 198 AlaLeu-AlaArgAlaProAspGlnProProSerProAspLeuProGluAspPro----- 215
QY 344 TCTGTGAGGACCG----- 370
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 216 -----SerProSerProProArgSerProAlaArgGlyGlnGlyProGlyArg 233
QY 371 AGGGGCTGCGCGGTGACCG-----CGCGGGGTGCAAA 406
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 233 gAspArgGluAlaLeuArgProCysArgSerGluSerProSerLeuGlyArgGlyProAr 253
QY 407 TAGCCATCATGACCTTCAAAGA 428
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 253 gAlaHisProSerLeuGlnLys 260

RESULT 12
T13412
Hypothetical protein I33E12.4 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13412
R:Murphy, L.; Harris, D.; Barrell, B.
Submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: Z17668
A:Accession: T13412
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2342 <MUR>
A:Cross-references: EMBL:AL009192; NID:e1371524; PID:e1202150; PIDN:CAA15686.1
C:Genetics:
A:Cross-references: FlyBase:FBgn0000667
A:Introns: 1161/3; 1205/1; 1283/1; 1432/3; 1489/2; 1912/3
A>Note: EG:I33E12.4

Alignment Scores:
Pred. No.: 1.19 Length: 2342
Score: 92.50 Matches: 37
Percent Similarity: 40.74% Conservative: 18
Best Local Similarity: 27.41% Mismatches: 41
Query Match: 8.40% Indels: 39
DB: 2 Gaps: 5

US-09-966-880A-7_COPY_80_676 (1-597) x T13412 (1-2342)

QY 173 AATTGCTCTTCCTCGCTACATCTCGACTGGACTAGACCTGGCCGCTGCTACCGCG 232
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1568 AsnValGlyAlaAspAlaAlaGlyAlaAlaGlyThrGlyThrValAlaAlaSerProGly 1587
QY 233 TCACCTGGTTACCTCCCTGGAGCCCTGCTAGACTGTGCCGACATGTGGCCGACTTC 292
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1588 GlySerGlySer-----GlyAla-----Asp-IleSerAsnValL 1599
QY 293 TGCAGAGGAACCCCAACCTCAGTCTGAGG---ATCTTCACCGCGCGC----- 336
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1599 euArgGlyAsnProAsnIleSerMetArgGluLeuPheHisGlyGluGluLeuGly 1619
QY 337 -----CTCTACTTCTGTGAGGACCGCAGCTGAGCCGAGGGCTGCGCGC 382
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1619 alGlnPheLysValProPheGlyCysSerSerGlnArgThrProGluGlyTrpThrA 1639
QY 383 GCCTGCACCGCCCGGGTGCAATAGCCATCATGACCTTCAAGATTATTTTACTGCT 442
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1639 rgVal----- 1640
```



```

QY 353 TCCTCACAAGTAGAGCGCGGTGAAGATCCTCAGACTGAGGTGGGTTCCTCCG 294
Db : : : : : ||| : : : |||
767 ProAlaGluLysSerArgSerArgArg-----ArgArgSerAlaSerSerProArg 784
QY 293 AGAAGTCGGCCACATGTCGGCCACAGTCG-----TAGCAGGGGCTCCAGGAG 246
Db ||| : : : : ||| : : : |||
785 ThrLysThrThrSerArgArgGlyArgSerProSerProLysProArgGlyLeuGlnArg 804
QY 246 ----- 246
Db
805 SerArgSerArgSerArgArgGluLysThrArgThrThrArgArgArgAspArgSergly 824
QY 245 GTGAACCCAGGTGACGGGTAGCAGCGGCCAGGCTCTAGGTCCAGTCCGAGATGTAGCGG 186
Db : : : ||| ||| : : : ||| ||| ||| : : : ||| : : : |||
825 SerSerGlnSerThrSerArgArgArgGlnArgSerArgSerArgSerArgValThrArg 844
QY 185 AGGAAGACCAATTCCACAGTGGCAGCGTCTTATTTCGGAAGATAACCAAGTCCAGTGAA 126
Db ||| ||| ||| ||| : : : |||
845 Arg-Arg-----ArgGlyGlySerGlyTyr-----HisSerArgSerProAla-- 858
QY 125 AGGATGTAGCAGTGTACCGCCTCTTCACTAGTACACAGGTAGGTCTCAGCGCCACCC 66
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
859 -ArgGlnGluSerSerArgThrSerSerArgArgArgArgGlyArgSerArgThrProPr 878
QY 65 T 65
Db 878 O 878

```

Search completed: June 14, 2003, 18:56:35  
Job time : 24.1832 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 14, 2003, 18:11:13 ; Search time 7.48841 Seconds

(without alignments)  
6613.252 Million cell updates/sec.

Title: US-09-966-880A-7\_COPY\_80\_676

Perfect score: 1101

Sequence: 1 atggcagcctcttgatgaa.....ttctactcttggactttga 597

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+n2p.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09966880/runat\_14062003\_175524\_10304/app\_query.fasta\_1.9493  
-DB=SwissProt\_40 -QFMT=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blsum62 -TRANS=human40.cd1  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09966880 @CGN\_1.1.196 @runat\_14062003\_175524\_10304 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query %	Match Length	ID	Description
1	364	33.1	382	1 PHB3_HUMAN	Q9uh17 homo sapien
2	363.5	33.0	199	1 PHB1_HUMAN	P31941 homo sapien
3	339	30.8	190	1 PHB2_HUMAN	Q9ue74 homo sapien
4	225	20.4	236	1 ABME_RABIT	F47855 oryctolagus
5	216.5	19.7	236	1 ABME_HUMAN	P41238 homo sapien
6	211	19.2	229	1 ABME_MOUSE	P51908 mus musculus
7	198	18.0	229	1 ABME_RAT	P38483 rattus norv
8	97.5	8.9	369	1 CT19_YEAST	Q02732 saccharomyc
9	94	9.1	504	1 ATIN_HSVBP	P30020 bovine herp
10	86.5	8.4	467	1 VU2_HPV24	P50770 human papil
11	86	7.8	522	1 NU62_HUMAN	P37198 homo sapien
12	85	7.7	503	1 MATK_KUNBA	Q9tkc0 kunzea baxt
13	85	7.7	2426	1 SON_HUMAN	P18583 homo sapien
14	84.5	7.7	525	1 NU62_RAT	P17955 rattus norv
15	84	8.1	2404	1 SON_MOUSE	Q9qx47 mus musculus
16	83.5	7.6	383	1 VGLZ_HSVKEK	P32515 equine herp
17	81.5	7.4	1185	1 DRPL_HUMAN	P54259 homo sapien
18	81	7.8	719	1 DEND_RAT	P50617 rattus norv

19	80.5	7.3	442	1 SDC3_MOUSE	Q64519 mus musculus
20	80.5	7.3	670	1 VG50_HSV11	Q00130 ictaluriid h
21	80.5	7.3	1446	1 IE18_PRVKA	P33479 pseudorabie
22	80.5	7.8	1758	1 CA14_CAEEL	P17139 caenorhabdi
23	80	7.3	546	1 AAAS_HUMAN	Q9arg9 homo sapien
24	79.5	7.2	797	1 VGLX_HSVBB	P28968 equine herp
25	78.5	7.1	275	1 HXAS_HETFR	Q9ia23 heterodontu
26	78.5	7.1	817	1 YG4B_YEAST	P46951 saccharomyc
27	78.5	7.6	1888	1 CA1E_CHICK	P32018 gallus gall
28	78	7.1	193	1 CIT1_HUMAN	Q99966 homo sapien
29	78	7.5	498	1 VE2_HPV08	P08422 human papil
30	78	7.1	1255	1 CD45_RAT	P04157 rattus norv
31	77.5	7.0	402	1 VGLD_PVRRI	P07645 pseudorabie
32	77.5	7.0	1367	1 AMYH_YEAST	P08640 saccharomyc
33	77.5	7.5	1403	1 CIC_DROME	Q9ulh0 drosophila
34	77.5	7.5	3119	1 CA1C_MOUSE	Q60847 mus musculus
35	77	7.0	525	1 NAB2_MOUSE	Q61127 mus musculus
36	77	7.0	1829	1 RW1_MOUSE	Q70472 mus musculus
37	76.5	6.9	642	1 NPT2_RABIT	Q28620 o renal sod
38	76.5	6.9	815	1 MK07_HUMAN	Q13164 homo sapien
39	76.5	7.4	1707	1 CA24_MOUSE	P08122 mus musculus
40	76	7.4	288	1 TR2B_HUMAN	Q15815 homo sapien
41	76	7.4	303	1 SPRI_ARATH	Q22315 arabidopsis
42	76	6.9	388	1 AZAB_ORYAF	Q19032 orycteropu
43	76	6.9	526	1 NU62_MOUSE	Q63850 mus musculus
44	76	6.9	662	1 MUC1_XENLA	Q05049 xenopus lae
45	76	6.9	2774	1 MAPA_RAT	P34926 rattus norv

#### ALIGNMENTS

RESULT 1

PHB3_HUMAN					
ID	PHB3_HUMAN	STANDARD;	PRT;	382 AA.	
AC	Q9UH17; 095618;				
DT	16-OCT-2001 (Rel. 40, Created)				
DE	16-OCT-2001 (Rel. 40, Last sequence update)				
DE	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Phorbolin 3 (APOBEC1-like).				
GN	APOBEC1L.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=20057165; PubMed=10591208;				
RA	Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,				
RA	Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,				
RA	Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,				
RA	Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,				
RA	Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,				
RA	Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,				
RA	Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,				
RA	Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,				
RA	Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,				
RA	Gilbert J.G.R., Goward M.E., Grafton D.V., Griffiths M.N.D., Hall C.,				
RA	Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,				
RA	Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,				
RA	Laird G.C., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,				
RA	Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,				
RA	Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,				
RA	Odeh C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,				
RA	Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,				
RA	Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,				
RA	Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,				
RA	Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,				
RA	Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,				
RA	Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,				
RA	Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,				
RA	Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,				
RA	Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,				
RA	Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,				





```

QY      424  AAAGATTATTTTACTGCTGGAATACCTTTGTA 456
Db      147  SerGluTyrCysTyrCysTrpGluAsnPheVal 157
          ::::: ||||| |||||
RESULT 5
ABME_HUMAN
ID      ABME_HUMAN      STANDARD;      PRT;      236 AA.
AC      P41238;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1996 (Rel. 33, Last sequence update)
DT      30-MAY-2000 (Rel. 39, Last annotation update)
DE      Apolipoprotein B mRNA editing protein (APOBEC-1) (HEPR).
GN      APOBEC1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Intestine;
RX      MEDLINE=942268910; PubMed=8208612;
RA      Hadjiadaplou C., Giannoni F., Funahashi T., Skarosi S.F.,
RA      Davidson N.O.;
RT      "Molecular cloning of a human small intestinal apolipoprotein B mRNA
RT      editing protein.";
RL      Nucleic Acids Res. 22:1874-1879(1994).
RN      [2]
RP      SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC      TISSUE=Small intestine;
RX      MEDLINE=94359963; PubMed=8078915;
RA      Lau P.P., Zhu H.-J., Baldini A., Charnsangavej C., Chan L.;
RT      "Dimeric structure of a human apolipoprotein B mRNA editing protein
RT      and cloning and chromosomal localization of its gene.";
RL      Proc. Natl. Acad. Sci. U.S.A. 91:8522-8526(1994).
RN      [3]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Peripheral blood leukocytes;
RX      MEDLINE=98140126; PubMed=9479499;
RA      Fujino T., Navaratnam N., Scott J.;
RT      "Human apolipoprotein B RNA editing deaminase gene (APOBEC1).";
RL      Genomics 47:266-275(1998).
CC      -1- FUNCTION: RESPONSIBLE FOR THE POSTTRANSCRIPTIONAL EDITING OF
CC      A CAA CODON FOR GLN TO A UAA CODON FOR STOP IN THE APOB MRNA.
CC      -1- COFACTOR: ZINC (BY SIMILARITY).
CC      -1- SUBUNIT: HOMODIMER.
CC      -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN THE SMALL INTESTINE.
CC      -1- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC      FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; L25877; AAA6766.1; -
EMBL; L26234; AAA64230.1; -
EMBL; AB009426; BAA23882.1; -
EMBL; AB009422; BAA23882.1; JOINED.
EMBL; AB009423; BAA23882.1; JOINED.
EMBL; AB009424; BAA23882.1; JOINED.
EMBL; AB009425; BAA23882.1; JOINED.
PIR; S45253; S45253.
Genew; HGNC:604; APOBEC1.
MIM; 600130; -
InterPro; IPR002125; dCMP/cyt_deam.
Pfam; PF00363; dCMP_cyt_deam; 1.
PROSITE; PS00903; CYT_DCMP_DEAMINASES; 1.
Kw mRNA processing; Hydrolase; zinc.
FT METAL 61 ZINC (BY SIMILARITY).

```

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; L25877; AAA6766.1; -;  
 DR EMBL; L26234; AAA64230.1; -;  
 DR EMBL; AB009426; BAA23882.1; -;  
 DR EMBL; AB009422; BAA23882.1; JOINED.  
 DR EMBL; AB009423; BAA23882.1; JOINED.  
 DR EMBL; AB009424; BAA23882.1; JOINED.  
 DR EMBL; AB009425; BAA23882.1; JOINED.  
 DR PIR; S45253; S45253.  
 DR Genew; HGNC:604; APOBEC1.  
 DR MIM; 600130; -;  
 DR InterPro; IPR002125; dCMP/cyt\_deam.  
 DR Pfam; PF00363; dCMP\_cyt\_deam; 1.  
 DR PROSITE; PS00903; CYT\_DCMP\_DEAMINASES; 1.  
 DR mRNA processing; hydrolase; zinc.  
 DR KEGG 61  
 DR METAL 61  
 CC ZINC (BY SIMILARITY).  
 -----





## Alignment Scores:

Pred. No.: 0.412 Length: 369  
 Score: 97.50 Matches: 50  
 Percent Similarity: 37.33% Conservative: 34  
 Best Local Similarity: 22.22% Mismatches: 70  
 Query Match: 8.86% Indels: 71  
 DB: 1 Gaps: 11

US-09-966-880A-7\_COPY\_80\_676 (1-597) x CT19\_YEAST (1-369)

```

QY 4 GACAGCCTTCTGATGACCGGAGGAGTTCTTTTACCAATTCAAAATGTCGCTGGCT 63
DB 62 AspAlaLeuThrArgArgAsnThrLeuLeuGlnGluLe 75
QY 64 AAGGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGGGGTGACAGTGTACATCC 123
DB 76 -----GlnThrTyrGlnAsnIleLeuMetLysGluAsnAsnSerLysThr--- 90
QY 124 TTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGGTGCCACGTGGAATGCTC--- 180
DB 91 -----LysAsnGly-----AspIleLeuGln 97
QY 181 -----TTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGC--- 219
DB 98 AsnAspIleThrGlnAspPheLeuAsnLeuIleSerIleSerSerSerAsnProAsnSer 117
QY 220 -----CCCTGCTACCGCGTCCACCTGGTTCACCTCCTGGAGC----- 255
DB 118 AlaIleSerAspArgLysArgValGluArgIleAsnGlyLeuThrAsnLeuGlnLysGlu 137
QY 256 -----CCCTGCTACGACTGTGCCGACATGTGGCGACTTTCTCGGAGGAACCCCAAC 309
DB 138 LeuValThrLysTyrAspThrLeuProLeuLeu-----Asn 149
QY 310 CTCAGTCTGAGATCTTACCGCGCGCTCTACTTCTGTGAGACCGCAAGGCTGAGCCC 369
DB 150 MetAsnLeuLeuSerTyrLeuArgAspHisThrTyrProHisLeuGlnValSerVal 169
QY 370 GAGGGCTCGCGGCTGCACCGCGCGGTGCAATAGCCATCATGACCTTCAAGAT 429
DB 170 GlnSerArgAspArgValHisAsnAspGlyIleGluValLeuValAsnIleLys--- 188
QY 430 TATTTTACTGCTGGAATACT-----TTGTAGAA 459
DB 189 -----PheCysArgAsnThrMetAsnProPheGluIleGlnPheLysMetPheTyrLys 206
QY 460 AACCATGAAGAACTTCAAGCCTCGGAAGGCTGCAT---GAAATTCAGTTCGTCTC 516
DB 207 PheGluAspSerThrLeuLeuLysTyrGluIleLeuArgIleSerThrAsnValArgLeu 226
QY 517 -----TCCAGACAGCTTCGGCGCATCCTTTTCCCTGATGAG 555
DB 227 LysAlaLysGlnLeuLeuAlaThrArgAsnPheGlnLysCysLeuLeuSerLeuTyrGlu 246
QY 556 GTTGATGACTACGA 570
DB 247 PheAspLysIleLys 251
  
```

## RESULT 9

ATIN\_HSVBP  
 ID ATIN\_HSVBP STANDARD; PRT; 504 AA.  
 AC P30020;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-FEB-1994 (Rel. 28, Last annotation update)  
 DE Alpha trans-inducing protein (Alpha-TIF).  
 OS Bovine herpesvirus type 1 (strain p8-2).  
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
 OC Alphaherpesvirinae; Varicellovirus.  
 OX NCBI\_TaxID=10324;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93012995; PubMed=1327963;

RA Carpenter D.E., Misra V.;  
 RT \*Sequences of the bovine herpesvirus 1 homologue of herpes simplex  
 RL virus type-1 alpha-trans-inducing factor (UL48).";  
 Gene 119:259-263(1992).  
 CC -!- FUNCTION: RESPONSIBLE FOR TRANSCRIPTIONAL ACTIVATION OF IMMEDIATE  
 CC EARLY PROMOTERS (ALPHA GENES) (BY SIMILARITY).  
 CC -!- SIMILARITY: TO OTHER HERPESVIRUSES ALPHA TRANS-INDUCING PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; Z11610; CAA77682.1; -;  
 DR PIR; S24229; S24229.  
 DR PIR; JCI306; JCI306.  
 DR InterPro; IPR003174; Alpha\_TIF.  
 DR Pfam; PF02232; Alpha\_TIF; 1.  
 KW Transcription regulation; Trans-acting factor; DNA-binding.  
 SQ SEQUENCE 504 AA; 54028 MW; 0031B4B0E31FCD25 CRC64;

## Alignment Scores:

Pred. No.: 0.875 Length: 504  
 Score: 94.00 Matches: 44  
 Percent Similarity: 33.75% Conservative: 10  
 Best Local Similarity: 27.50% Mismatches: 41  
 Query Match: 9.09% Indels: 65  
 DB: 1 Gaps: 7

US-09-966-880A-7\_COPY\_80\_676 (1-597) x ATIN\_HSVBP (1-504)

```

QY 569 CGTAAGTCATCAACCTCATCAGGGGCAAA-----AGGATGGCG 531
DB 361 ArgSerGlySerThrArgThrArgGlyArgAlaAlaArgSerThrGlyArgLeuGln 380
QY 530 CGAAGCTGTCTGGAGAGACGAACGAATTTTCATGC-----AGCCCTCCCGAGCT 480
DB 381 ArgProCysGlyProArgArgArgAlaLysCysArgAlaThrProArgGlnArg 400
QY 479 TTGAAGACTTCTTTCATGTTTTCATCAAAAGATTATCCAGCAGTAAATAATCTTTGAAG 420
DB 401 LeuArg----- 402
QY 419 GTCATGATGGCTATTTCACCCCGCGGGTGCAGCCCGCCGCGCTCAGCC 360
DB 403 -----AlaArgGlyGluProArgHisThrSerGlySerGly 414
QY 359 TTGGGTCTCTCACAGAAGTAGAGCGCGGTGAAGATCCTCAGACTGAGGTGGGGTTC 300
DB 415 AlaPheSerGlnGly-----ArgArgProGlyArgValCysArgLeu-GlyTrpAlaCys 432
QY 299 CCTCGCAAGAGTCGGCCA-----CATGT 276
DB 432 LysAlaArgSerGlyProAlaArgGlyGlyProGlyProSerProValArgSerGlyLe 452
QY 275 CGGGCAGACGTGTAGCAGGGGCTCCAGGAGGTGAACACGAGTGCAGCGGTAGCGGCCA 216
DB 452 uGlyLeuSerArgAlaArgGlySerPro----- 461
QY 215 GGGTCTAGTCCAGTCCGAGATGTAGCGGAGGAGAGCAATTCACGTGGCAGCGCT 158
DB 462 -----GlyProGlyProAlaCys---GlyGlyProSerArgAlaArgGlyArg 477
  
```

RESULT 10  
 VE2\_HPV24  
 ID VE2\_HPV24 STANDARD; PRT; 467 AA.  
 AC P50770;  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)



1

NU62\_HUMAN STANDARD; PRT; 522 AA.  
P37198; Q9NSL1; Q96C43;  
01-OCT-1994 (Rel. 30, Created)  
15-JUN-2002 (Rel. 41, Last sequence update)  
15-JUN-2002 (Rel. 41, Last annotation update)  
Nuclear pore glycoprotein p62 (62 kDa nucleoporin).  
NUP62.  
Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92007939; PubMed=1915414;  
RA Carmo-Fonseca M., Kern H., Hurt E.C.;  
RT "Human nucleoporin p62 and the essential yeast nuclear pore protein  
NSP1 show sequence homology and a similar domain organization.";  
RL Eur. J. Cell Biol. 55:17-30(1991).  
[2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Bloecker H., Boecher M., Brandt P., Mewes H.-W., Weil B., Wiemann S.;  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
[3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin, and Urinary bladder;  
RA Strausberg R.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX.  
CC THE N-TERMINAL IS PROBABLY INVOLVED IN NUCLEOCYTOPLASMIC  
CC TRANSPORT. THE C-TERMINAL IS PROBABLY INVOLVED IN PROTEIN-PROTEIN  
CC INTERACTION VIA COILED-COIL FORMATION AND MAY FUNCTION IN  
CC ANCHORAGE OF P62 TO THE PORE COMPLEX.  
CC -!- SUBCELLULAR LOCATION: CENTRAL REGION OF THE NUCLEAR PORE, WITHIN  
CC THE TRANSPORTER. DURING MITOTIC CELL DIVISION, IT ASSOCIATES WITH  
CC THE POLES OF THE MITOTIC SPINDLE.  
CC -!- DOMAIN: CONTAINS F-X-F-G REPEATS.  
CC -!- PTM: O-GLYCOSYLATED. CONTAINS ABOUT 10 N-ACETYLGLUCOSAMINE SIDE  
CC CHAIN SITES PREDICTED FOR THE ENTIRE PROTEIN, AMONGST WHICH ONLY  
CC ONE IN THE C-TERMINAL  
CC -!- SIMILARITY: BELONGS TO THE NUP62 FAMILY.  
-----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; X59521; CAA41411.1; -  
DR EMBL; AL162061; CAB82399.1; -  
DR EMBL; BC003663; AAH03663.1; -  
DR EMBL; BC014842; AAH1842.1; -  
DR GlycoSuiteDB; P37198; -  
DR Genew; HGNC:8066; NUP62.  
DR MIM; 605815; -  
KW Nuclear protein; Transport; Glycoprotein; Coiled coil; Repeat;  
KW Polymorphism.  
FT DOMAIN 9 288 THR-RICH.  
FT DOMAIN 176 331 ALA-RICH.  
FT FT 1 295 15 X 9 AA APPROXIMATE REPEATS.  
FT REPEAT 1 9 1.  
FT REPEAT 13 21 2.  
FT REPEAT 29 37 3.  
FT REPEAT 39 47 4.  
FT REPEAT 57 65 5.  
FT REPEAT 71 79 6.  
FT REPEAT 85 93 7.  
FT REPEAT 111 119 8.  
FT REPEAT 137 145 9.  
FT REPEAT 155 163 10.  
FT REPEAT 168 176 11.



```
FT REPEAT 185 193 12.
FT REPEAT 234 242 13.
FT REPEAT 253 261 14.
FT REPEAT 287 295 15.
FT DOMAIN 328 458 COILED COIL (POTENTIAL).
FT CARBOHYD 373 373 O-LINKED (GLNAC) (BY SIMILARITY).
FT CARBOHYD 468 468 O-LINKED (GLNAC) (BY SIMILARITY).
FT VARIANT 233 233 A -> S (IN DBSNP:2290772).
/FTid=VAR_013467.
FT CONFLICT 283 283 T -> S (IN REF. 3: AAH14842).
FT CONFLICT 418 419 SG -> RA (IN REF. 1).
FT CONFLICT 431 431 E -> Q (IN REF. 1).
FT CONFLICT 507 507 E -> V (IN REF. 1).
SQ SEQUENCE 522 AA; 53268 MW; EAA3054F452719A2 CRC64;

Alignment Scores:
Pred. No.: 4.57 Length: 522
Score: 86.00 Matches: 56
Percent Similarity: 37.05% Conservative: 27
Best Local Similarity: 25.00% Mismatches: 89
Query Match: 7.81% Indels: 52
DB: 1 Gaps: 7

US-09-966-880A-7_COPY_80_676 (1-597) x NU62_HUMAN (1-522)
QY 35 TTTACCAATTCAAAATGTCCGCTGAAGGTCGGCTGAGACCTACCTGTGCTACG 94
||||| : : : : : |||
Db 190 PheThrProAlaThrProAlaAlaThrAlaGlyAlaThrGlnProAlaAlaProThr 209
QY 95 TAGTGAAGAGCGGTGACAGTGCATACCTTTTCACGTGGACTTTGGTTACTTCGCA--- 151
: : : : : |||
Db 210 ProThrAlaThrIleThrSerThrGlyPro-----SerLeuPheAlaSer 224
QY 152 ATAAGAGCGTCCGACGTGGAATGCTCTCCCTCCGCTACATCTCGAGCTGGACCTAG 211
||| ||||| ||| : : : : : |||
Db 225 IleAlaThrAlaProThrSerSerAlaThrThrGlyLeuSerLeuCysThrProValThr 244
QY 212 ACCCTGCGCGCTGCTACCGCG-----TCACCTGGTTCACCT 247
||| ||| ||||| : : : : : |||
Db 245 ThrAlaGlyAlaProThrAlaGlyThrGlnGlyPheSerLeuLysAlaProGlyAlaAla 264
QY 248 CTGGAGCCCTGCTAGGACTGTGCCGACATGTGGCGACTTTCTCGGAGGGAACCCCA 307
||| ||||| ||| |||
Db 265 SerGlyThrSerThrThrThrSerThrAlaAlaThrAlaThrThrThrThrSer 284
QY 308 ACCTCAGTCTGAGGATCTCA-----CCGCGC 334
: : : : : |||
Db 285 SerSerThrThrGlyPheAlaLeuAsnLeuLysProLeuAlaProAlaGlyIleProSer 304
QY 335 GCCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCGCGGCTGCGGCGGTGCACCGCG 394
: : : : : ||||| ||| ||||| ||||| |||
Db 305 AsnThrAlaAlaAlaValThrAlaProGlyPro-GlyAlaAlaAlaGlyAlaAlaAl 324
QY 395 CCGGGTGAATAAGCATCATGACCTTCAAGATTATTTTACTGCTGAATCTTTG 454
||| : : : : : |||
Db 324 aSerSerAlaMetThrTyrAlaGlnLeuGluSerLeu-IleAsnLysTrpSerLeuGluL 344
QY 455 TAGAAACCATGAAGAACTTTC-----AAAGCTGGGAA--- 489
: : : : : ||||| |||
Db 344 euGluaspGlnGluArgHisPheLeuGlnGlnAlaThrGlnValAsnAlaTrpaspArg 364
QY 490 -----GGGCTGCATGAA----- 501
|||
Db 364 hrLeuIleGluAsnGlyGluLysIleThrSerLeuHisArgGluValGluLysValLysL 384
QY 502 --AATTCAGTTCGCTCTCCAGACAGATTTCGGCGCATCTTTTCCCGCTGTATGAGTTG 559
: : : : : ||||| |||||
Db 384 euAspGlnLysArgLeuAspGlnGlnLeuAspPheIleLeuSerGlnGlnLysGluLeuG 404
QY 560 ATGACTTA 567
Db 404 luAspLeu 406
```

```
RESULT 12
MATK_KUNBA
* ID MATK_KUNBA STANDARD; PRT; 503 AA.
AC Q9PKC0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK.
OS Kunzea baxteri (Scarlet kunzea).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Myrtales; Myrtaceae; Kunzea.
OX NCBI_TaxID=106042;
RN [1]
RP SEQUENCE FROM N.A.
RA O'Brien M.M., Wilson P.G., Quinn C.J.;
RT "Molecular systematics of the Leptospermum suballiance (Myrtaceae).";
RL Aust. J. Bot. 48:621-628(2000).
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS.
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED
CC BY MITOCHONDRIAL INTRONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF104722; AAF05929.1; -.
DR InterPro; IPR000442; Intron_maturase2.
DR InterPro; IPR002866; MatK_N.
DR Pfam; PF01348; Intron_maturase2; 1.
DR Pfam; PF01824; MatK_N; 1.
KW Chloroplast; mRNA processing.
SQ SEQUENCE 503 AA; 60203 MW; B114A4704FCEA059 CRC64;

Alignment Scores:
Pred. No.: 5.59 Length: 503
Score: 85.00 Matches: 35
Percent Similarity: 41.43% Conservative: 23
Best Local Similarity: 25.00% Mismatches: 52
Query Match: 7.72% Indels: 30
DB: 1 Gaps: 8

US-09-966-880A-7_COPY_80_676 (1-597) x MATK_KUNBA (1-503)
QY 91 TACGTAGTGAAGCGGTGACAGTCTACATCTTTTCTACTGAC----- 135
: : : : : ||||| ||| ||||| : : :
Db 112 PheSerLeuGluArgLysGluIleAlaLysSerTyrAsnLeuArgSerIleHisSerIle 131
QY 136 TTTGGTTATCTTCGCAATAAG----- 156
||| : : : : : |||
Db 132 PheSerPheLeuGluAspLysPheThrHisLeuAspTyrValSerAspValLeuIlePro 151
QY 157 AACGGCTCCACGTGGAAATGCTCTTC-----CTCCGCTACATCTCGGACGTGGACCTA 210
||| : : : : : |||
Db 152 TyrHisIleHisLeuGluIleLeu**GlnThrLeuArgTyr-----TrpValLys 168
QY 211 GACCTGCGCGCTGCTACCGGTACCTGGTTTCACCTCC-----TGGAGCCCTGCTAC 264
||| : : : : : |||
Db 169 AspaLaserSerLeuHisLeuArgPhePheLeuHisAspTyrTrpAsnSerPheIle 188
QY 265 GACTGTGCGCGACATGTGGCGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATC 324
||| : : : : : ||||| ||| |||
Db 189 ThrProLysLysHisIleThrPhePheLeuLysGlyAsnProArgLeuPheLeu 208
QY 325 TTCACGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGGAGGGGCTGCGCGCG 384
```

Db 209 TyrAsnSer-----HisIleCysGluTyrGluTyrIlePhePro----PheLeuArgAsn 225  
Qy 385 -----CTGCACCGCGCGGGTCAATAATGACCATCATGACCTTCRAAGATTATTTTAC 438  
Db 226 GlnSerSerHisLeuArgSerThrSerSerGlyIlePhe\*\*GluArgIleTyrPheTyr 245

RESULT 13  
SON\_HUMAN  
ID SON\_HUMAN STANDARD: PRT: 2426 AA.  
AC Q9P072; Q95981; Q9UPY0; Q14120; Q14487; Q9UKP9; Q9H7B1; Q9P070;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE SON protein (SON3) (Negative regulatory element-binding protein) (NRE-  
binding protein) (DBP-5) (Bax antagonist selected in saccharomyces 1)  
DE (BASS1) (protein C21orf50).  
GN SON OR NREBP OR DBP5 OR C21ORF50 OR KIAA1019.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS A; B; C; D; E AND F).  
RX MEDLINE=21564202; PubMed=11707072;  
RA Raymond A., Friedli M., Neergaard Henriksen C., Chapot F.,  
RA Deutsch S., Ucla C., Rossier C., Lyle R., Guipponi M.,  
RA Antonarakis S.E.;  
RT "From PRDs and open reading frames to cDNA isolation: revisiting the  
RT human chromosome 21 transcription map.";  
RL Genomics 78:46-54(2001).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM G).  
RC TISSUE=Liver;  
RX MEDLINE=21316479; PubMed=11306577;  
RA Sun C.-T., Lo W.-Y., Wang I.-H., Lo Y.-H., Shiou S.-R., Lai C.-K.,  
RA Ting L.-P.;  
RT "Transcription repression of human hepatitis B virus genes by negative  
RT regulatory element-binding protein/SON.";  
RL J. Biol. Chem. 276:24059-24067(2001).  
RN [3]  
RP SEQUENCE OF 1-689 FROM N.A. (ISOFORM H).  
RC TISSUE=Placenta;  
RA Casadei R., Strippoli P., D'Addabbo P., Canaider S., Lenzi L.,  
RA Vitale L., Giannone S., Carinci P., Zannotti M.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
RN [4]  
RP SEQUENCE OF 1-130 FROM N.A.  
RC TISSUE=Smooth muscle;  
RA Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K.,  
RA Nakajima Y., Mizuno T., Morinaga M., Tanigami A., Fujiwara T., Ono T.,  
RA Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Ota T., Suzuki Y.,  
RA Obayashi M., Nishi T., Shibahara T., Tanaka T., Nakamura Y.,  
RA Isogai T., Sugano S.;  
RT "NEDO human cDNA sequencing project.";  
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
RN [5]  
RP SEQUENCE OF 1-114 FROM N.A.  
RC TISSUE=Blood;  
RA Ye M., Zhang Q.H., Zhou J., Shen Y., Wu X.Y., Guan Z.Q., Wang L.,  
RA Fan H.Y., Mao Y.F., Dai M., Huang Q.H., Chen S.J., Chen Z.;  
RT "Human partial CDS from cd34+ stem cells.";  
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.  
RN [6]  
RP SEQUENCE OF 437-2426 FROM N.A. (ISOFORM B).  
RC TISSUE=Brain;  
RX MEDLINE=99397452; PubMed=10470851;  
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,  
RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
RT "Prediction of the coding sequences of unidentified human genes. XIV.  
RT The complete sequences of 100 new cDNA clones from brain which code  
RT for large proteins in vitro.";

RL DNA Res. 6:197-205(1999).  
RN [7]  
RP SEQUENCE OF 554-2426 FROM N.A. (ISOFORM A).  
RX MEDLINE=92049296; PubMed=1944255;  
RA Chumakov I.M., Berdichevskii F.B., Sokolova N.V., Reznikov M.V.,  
RA Prasolov V.S.;  
RT "Identification of a protein product of a novel human gene SON and  
RT the biological effect upon administering a changed form of this gene  
RT into mammalian cells.";  
RL Mol. Biol. (Mosk) 25:731-740(1991).  
RN [8]  
RP SEQUENCE OF 709-1079 FROM N.A. (ISOFORM I).  
RC TISSUE=Placenta;  
RX MEDLINE=93062885; PubMed=1435774;  
RA Bliskovskii V.V., Kirillov A.V., Zakhariev V.M., Chumakov I.M.;  
RT "The human son gene: the large and small transcripts contains various  
RT 5'-terminal sequences.";  
RL Mol. Biol. (Mosk) 26:807-812(1992).  
RN [9]  
RP SEQUENCE OF 1009-1131 FROM N.A.  
RC TISSUE=Placenta;  
RX MEDLINE=93062884; PubMed=1435773;  
RA Bliskovskii V.V., Berdichevskii F.B., Tkachenko A.V., Belova M.E.,  
RA Chumakov I.M.;  
RT "Coding part of the son gene small transcript contains four areas of  
RT complete tandem repeats.";  
RL Mol. Biol. (Mosk) 26:793-806(1992).  
RN [10]  
RP SEQUENCE OF 1145-2426 FROM N.A. (ISOFORM F).  
RX MEDLINE=93048367; PubMed=1424986;  
RA Mattioni T., Hume C.R., Konigorski S., Hayes P., Osterweil Z.,  
RA Lee J.S.;  
RT "A cDNA clone for a novel nuclear protein with DNA binding  
RT activity.";  
RL Chromosoma 101:618-624(1992).  
RN [11]  
RP SEQUENCE OF 1692-2175 FROM N.A. (ISOFORM A).  
RX MEDLINE=89039788; PubMed=3054499;  
RA Berdichevskii F.B., Chumakov I.M., Kiselev L.L.;  
RT "Decoding of the primary structure of the son3 region in human  
RT genome: identification of a new protein with unusual structure and  
RT homology with DNA-binding proteins.";  
RL Mol. Biol. (Mosk) 22:794-801(1988).  
RN [12]  
RP SEQUENCE OF 1939-2426 FROM N.A. (ISOFORM J).  
RC TISSUE=Cerebellum;  
RX MEDLINE=99439804; PubMed=10509013;  
RA Greenhalf W., Lee J., Chaudhuri B.;  
RT "A selection system for human apoptosis inhibitors using yeast.";  
RL Yeast 15:1307-1321(1999).  
RN [13]  
RP FUNCTION: Represses hepatitis B virus (HBV) core promoter activity  
RP and transcription of HBV genes and production of HBV virions.  
CC Binds to the consensus DNA sequence: 5'-GA[GT]AN[CG][AG]CC-3'.  
CC Might protect cells from apoptosis. Might be involved in pre-mRNA  
CC splicing (By similarity).  
CC -|- SUBCELLULAR LOCATION: Nuclear with a speckled distribution.  
CC -|- ALTERNATIVE PRODUCTS: 10 isoforms: A, B, C, D, E, F (shown here),  
CC G, H, I and J; may be produced by alternative splicing  
CC -|- TISSUE SPECIFICITY: Widely expressed, with the higher expression  
CC seen in leukocyte and heart.  
CC -|- DOMAIN: Contains 8 types of repeats which are distributed in 3  
CC regions.  
CC -|- MISCELLANEOUS: Colocalizes with the pre-mRNA splicing factor  
CC SFRS2/SC-35.  
CC -|- SIMILARITY: CONTAINS 1 G-PATCH DOMAIN.  
CC -|- SIMILARITY: CONTAINS 1 DRBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.  
CC -|- CAUTION: ISOFORM A SEQUENCE FROM REF.7 DIFFERS FROM THAT SHOWN  
CC DUE TO A FRAMESHIFT.  
CC -|- CAUTION: ISOFORM F SEQUENCE FROM REF.10 DIFFERS FROM THAT SHOWN  
CC DUE TO A FRAMESHIFT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation



```

[4] CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=95151751; PubMed=7849028;
RA Lubas W.A., Smith M., Starr C.M., Hanover J.A.;
RT "Analysis of nuclear pore protein p62 glycosylation.";
RL Biochemistry 34:1686-1694(1995).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF THE NUCLEAR PORE COMPLEX.
CC THE N-TERMINAL IS PROBABLY INVOLVED IN NUCLEOCYTOPLASMIC
CC TRANSPORT. THE C-TERMINAL IS PROBABLY INVOLVED IN PROTEIN-PROTEIN
CC INTERACTION VIA COILED-COIL FORMATION AND MAY FUNCTION IN
CC ANCHORAGE OF P62 TO THE PORE COMPLEX.
CC -!- SUBCELLULAR LOCATION: CENTRAL REGION OF THE NUCLEAR PORE, WITHIN
CC THE TRANSPORTER. DURING MITOTIC CELL DIVISION, IT ASSOCIATES WITH
CC THE POLES OF THE MITOTIC SPINDLE.
CC -!- DOMAIN: CONTAINS F-X-F-G REPEATS.
CC -!- SIMILARITY: BELONGS TO THE NUP62 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04143; AAA60741.1; -.
DR EMBL; X52583; CAA36813.1; -.
DR EMBL; M62992; AAA41789.1; -.
DR PIR; A31762; A31762.
DR PIR; A35596; A35596.
DR GlycoSuiteDB; PI7955; -.
KW Nuclear protein; Transpore; Glycoprotein; Coiled coil; Repeat.
FT DOMAIN 17 294 THR-RICH.
FT DOMAIN 105 269 ALA-RICH.
FT DOMAIN 331 461 COILED COIL (POTENTIAL).
FT DOMAIN 271 284 POLY-THR.
FT DOMAIN 291 294 POLY-THR.
FT CARBOHYD 376 376 O-LINKED (GLCNAC).
FT CARBOHYD 471 471 O-LINKED (GLCNAC).
FT CONFLICT 371 372 NG -> FR (IN REF. 3).
SQ SEQUENCE 525 AA; 53396 MW; 6F02F6BF6C0816E CRC64;

Alignment Scores:
Pred. No.: 6.22 Length: 525
Score: 84.50 Matches: 57
Percent Similarity: 37.87% Conservative: 32
Best Local Similarity: 24.26% Mismatches: 79
Query Match: 7.67% Indels: 67
DB: 1 Gaps: 10

```

**Alignment Scores:**

Pred. No.:	6.22	Length:	525
Score:	84.50	Matches:	37
Percent Similarity:	37.87%	Conservative:	52
Best Local Similarity:	24.26%	Mismatches:	79
Query Match:	7.67%	Indels:	67
DB:	1	Gaps:	10

US-09-966-880A-7\_COPY\_80\_676 (1-597) x NU62\_RAT (1-525)

Qy	26	GGAAGTTCTTTACCAATTCAAAATGTCC-----	-CCTGGCTAAGGGTC	70
Db	189			
		GlySerProPheThrAlaAlaThrLeuAlaThrThrAlaGlyAlaThrGlnProAla		208
Qy	71	GGCGTGAGACCTACCTGTCTACGTAGTCAAGAGCGGTGCACAGTGCTA-----		118
Db	209			
		AlaAlaThrProThrAlaAlaThrThrSerAlaGlySerThrLeuPheAlaSerIleAla		228
Qy	119	--CATCTTTTCTACTGGCTTTGGTTATPTCTGC AATAAGACCGCTGCCACGTGGAAAT		175
Db	229			
		AlaAlaProAlaSerSerSerThrThrValLeuSerLeuSerAlaProAlaThr-----		246
Qy	176	TGCTCTTCTCCGCTACATCTCGGACTGGGACCTAGACCTCGCGCTGCTACCGCGTCA		235
Db	247	-----ThrAlaAlaThrProThrAlaGly-----	ThrLeuGlyPheSerLeuIlysAla	262
Qy	236	CCTGGTTTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCGTC		295
Db	263	ProGlyAlaAlaProGlyAlaSerThrThrSer-----		
		-----ThrThrThr		276
Qy	296	GAGGAAACCCCAACTCAGTCTGAGATCTTTCACCGCGCGGCTCTACTCTCTGTGAGGA-----		353





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 14, 2003, 18:12:03 ; Search time 36.6063 Seconds  
(without alignments)  
6720.711 Million cell updates/sec

Title: US-09-966-880a-7\_COPY\_80\_676

Perfect score: 1101

Sequence: 1 atggacagccttgatgaa.....ttctacttgggactttga 597

Scoring table: BLOSUM62  
Ygapop 10.0, Ygapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09966880/runat\_14062003\_175524\_10316/app\_query.fasta\_1.9493  
-DB=SPTRMBL\_21 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS-human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09966880@cgn\_1.1.978 @runat\_14062003\_175524\_10316 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOPEXT=10 -XGAPOPEXT=0.5 -FGAPOPEXT=6  
-FGAPEXT=7 -YGAPOPEXT=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phase.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_rviro.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaea.\*

Pred, No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1086	98.6	198 4 Q9GZX7	Q9gzx7 homo sapien

2	1008	91.6	198 11 Q9WVE0	Q9wve0 mus musculu
3	390	35.4	384 4 Q9HC16	Q9hc16 homo sapien
4	369.5	33.6	204 4 Q9Y555	Q9y555 homo sapien
5	367.5	33.4	190 4 Q96F12	Q96f12 homo sapien
6	354	32.2	294 4 Q9NRW3	Q9nrw3 homo sapien
7	349	31.7	429 11 Q99J72	Q99j72 mus musculu
8	336.5	30.6	386 4 Q96AK3	Q96ak3 homo sapien
9	315.5	28.7	139 4 Q9Y4V1	Q9y4v1 homo sapien
10	278	25.2	224 11 Q9WV35	Q9wv35 mus musculu
11	274	24.9	224 4 Q9Y235	Q9y235 homo sapien
12	270.5	24.6	121 4 Q9Y553	Q9y553 homo sapien
13	257.5	23.4	103 4 Q9Y550	Q9y550 homo sapien
14	251	22.8	111 4 Q9UHI8	Q9uhi8 homo sapien
15	240.5	21.8	121 4 Q9Y554	Q9y554 homo sapien
16	238.5	21.7	131 4 Q9Y530	Q9y530 homo sapien
17	238.5	21.7	245 4 Q9Y549	Q9y549 homo sapien
18	219	19.9	229 11 Q9EQP0	Q9eqp0 mesocricetu
19	218.5	19.8	236 4 Q9UM71	Q9um71 homo sapien
20	213.5	19.4	235 6 Q9TUI7	Q9tui7 monodelphis
21	211	19.2	229 11 Q99L67	Q99l67 mus musculu
22	198	18.0	127 4 Q9H9H8	Q9h9h8 homo sapien
23	182.5	16.6	77 4 Q9Y551	Q9y551 homo sapien
24	179	16.3	132 4 Q9UE64	Q9ue64 homo sapien
25	102	9.3	720 5 Q9W293	Q9w293 drosophila
26	94.5	8.6	568 11 Q9EPR8	Q9epr8 mus musculu
27	93.5	8.5	568 11 Q9ESJ1	Q9esj1 mus musculu
28	93	8.4	277 11 Q06477	Q06477 mus musculu
29	92.5	8.4	2342 5 Q46048	Q46048 drosophila
30	91.5	8.3	839 10 Q9SN46	Q9sn46 arabidopsis
31	91	8.2	249 2 Q8VPM4	Q8vpm4 micrococcus
32	90	8.7	591 4 Q9UOL3	Q9uol3 homo sapien
33	89.5	8.1	629 12 Q9DWH3	Q9dwh3 rat cytomeg
34	89	8.6	429 4 Q00112	Q00112 homo sapien
35	88.5	8.0	1493 11 P81128	P81128 rattus norv
36	87.5	7.9	464 5 Q9NAX4	Q9nax4 dictyostell
37	87.5	8.5	583 4 Q9BT33	Q9bt33 homo sapien
38	87.5	8.5	585 4 Q9BWS2	Q9bws2 homo sapien
39	87.5	8.5	1275 4 Q9UQ36	Q9uq36 homo sapien
40	87.5	8.5	1783 4 Q15038	Q15038 homo sapien
41	87.5	8.5	1791 4 Q60382	Q60382 homo sapien
42	87.5	8.5	2296 4 Q9UHA8	Q9uha8 homo sapien
43	87.5	8.5	2752 4 Q9UQ35	Q9uq35 homo sapien
44	87	8.4	513 5 Q9VL71	Q9vl71 drosophila
45	86.5	7.9	209 10 Q9LWP8	Q9lwp8 oryza sativ

## ALIGNMENTS

RESULT 1

Q9GZX7 Q9GZX7 PRELIMINARY; PRT; 198 AA.  
ID Q9GZX7 Q9GZX7;  
AC Q9GZX7;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)  
DE Activation-Induced cytidine deaminase.  
GN AID.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20408890; PubMed=10950930;  
RA Muto T., Muramatsu M., Taniwaki M., Kinoshita K., Honjo T.;  
RT "Isolation, tissue distribution and chromosomal localization of the  
RL human activation-induced cytidine deaminase (hAID) gene.";  
RN Genomics 68:85-88(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20460541; PubMed=11007475;  
RA Revy P., Muto T., Levy Y., Geissmann F., Plebani A., Sanal O.,  
RA Catalan N., Forveille M., Dufourcq-Lagelouse R., Gennery A.,





## RESULT 3

Q9HC16 ID Q9HC16 PRELIMINARY; PRT; 384 AA.  
AC Q9HC16; 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE MDS019 (Phorbol-like protein MDS019).  
GN MDS019.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Huang C., Qian B., Tu Y., Gu W., Wang Y., Han Z., Chen Z.;  
RT "Novel genes expressed in hematopoietic stem/progenitor cells from  
RT Myelodysplastic Syndromes patient."  
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Straussberg R.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF182420; AAG14956.1; -.  
DR EMBL; BC024268; AAH24268.1; -.  
DR InterPro; IPR002125; dCMP/cyt\_deam.  
DR PROSITE; PS00903; CYT\_DCMP\_DEAMINASES; UNKNOWN\_1.  
SQ SEQUENCE 384 AA; 46408 MW; 60525DC3B7D903D6 CRC64;

## Alignment Scores:

Pred. No.: 1.09e-30 Length: 384  
Score: 390.00 Matches: 83  
Percent Similarity: 61.62% Conservative: 31  
Best Local Similarity: 44.86% Mismatches: 59  
Query Match: 35.42% Indels: 12  
DB: 4 Gaps: 3

US-09-966-880A-7\_COPY\_80\_676 (1-597) x Q9HC16 (1-384)

QY 16 ATGACCGGAGGAGTTCTTTACCAATTCGCGTGGCGTAAAGGTCGGCGT 75  
DB 197 MetAspProThrPheThrPheAsnGluProTrpValArgGlyArgHis 216  
QY 76 GAGACCTACTGTGCTACCTAGTGAAGAGCGCGTACACCTCTTTTACACGGAC 135  
DB 217 GluThrTyrLeuCysTyrGluValGluArgMethHisAsnAspThrTrpValLeuLeuAsn 236  
QY 136 -----TTTGGTTATCTTCGCAATAAG-----AACGGCTGC 165  
DB 237 GluArgArgGlyPheLeuCysAsnGlnAlaProHisLysHisGlyPheLeuGluGlyArg 256  
QY 166 CAGCTGGAAATGCTCTTCCTCCGCTACATCTCGGACTGGACCTAGACCTGGCGCTGC 225  
DB 257 HisAlaGluLeuCysPheLeuAspValIleProPheTrpLysLeuAspLeuAspGlnAsp 276  
QY 226 TACCGCGTCACTGGTTCACTCTGAGCCCTGCTACGACGTGCGCGACATGGCC 285  
DB 277 TyrArgValThrCysPheThrSerTrpSerProCysPheSerCysAlaGlnGluMetAla 296  
QY 286 GACTTTCTCGGAGGGAACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTTC 345  
DB 297 LysPheIleSerLysAsnLysHisValSerLeuCysIlePheThrAlaArgIleTyr--- 315  
QY 346 TGTGAGGACCGCAAGCTGAGCCCGAGGGGCTCGCGGCTGACCGCGCGGGGTGCA 405  
DB 316 ---AspAspGlnGlyArgCysGlnGluGlyLeuArgThrLeuAlaGluAlaGlyAlaLys 334  
QY 406 ATAGGCATCATGACCTTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAACCAT 465  
DB 335 IleSerIleMetThrTyrSerGluPheLysHisCysTrpAspThrPheValAspHisGln 354  
QY 466 GAAAGAACTTTCAAAGCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCTCCAGACAG 525

Db 355 GlyCysProPheGlnProTrpAspGlyLeuAspGluHisSerGlnAspLeuSerGlyArg 374  
QY 526 CTTTCGGCGCATCTT 540  
Db 375 LeuArgAlaIleLeu 379  
RESULT 4  
Q9Y555 ID Q9Y555 PRELIMINARY; PRT; 204 AA.  
AC Q9Y555; 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE BK150C2.3 (Putative novel protein similar to APOBEC1 (Apolipoprotein B  
DE mRNA editing protein) and Phorbolin) (Fragment).  
GN BK150C2.3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Matthews L.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL022318; CAB45271.1; -.  
DR InterPro; IPR002125; dCMP/cyt\_deam.  
DR PROSITE; PS00903; CYT\_DCMP\_DEAMINASES; UNKNOWN\_1.  
KW Lipoprotein.  
FT NON\_TER 1  
SQ SEQUENCE 204 AA; 24285 MW; 79C656F580A40554 CRC64;

## Alignment Scores:

Pred. No.: 1.13e-28 Length: 204  
Score: 369.50 Matches: 79  
Percent Similarity: 58.52% Conservative: 24  
Best Local Similarity: 44.89% Mismatches: 64  
Query Match: 33.56% Indels: 9  
DB: 4 Gaps: 4

US-09-966-880A-7\_COPY\_80\_676 (1-597) x Q9Y555 (1-204)

QY 31 TTTCTTTTACCAATTCAAAAATGTCGCGTGGCGTAAAGGTCGGCGTACCTGTGC 90  
DB 31 PheThrPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuGly 50  
QY 91 TAGCTAGTGAAG-----AGCGTGACAGTGCTACATCTTTTCTCAGTGGCTTGTAT 144  
DB 51 PheThrValGluGlyIleLysArgSerValValSerTrpLysThr-----GlyVal 68  
QY 145 CTTTCGCAAT-----AAGAACGGCTGCCACGTGGGAATTCCTCTCTCCGCTAC 192  
DB 69 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 88  
QY 193 ATCTCGGCTGGGACCTAGACCTGCGCGTGTACCGCGTACCTGTTTACCTCCCTGG 252  
DB 89 PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTyrThrSerTrp 108  
QY 253 AGCCCTGCTACGACTGTGCGCGACATGTGCGCGACTTCTGCGGAGGAACCCCAACCTC 312  
DB 109 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 128  
QY 313 AGTCTGAGGATCTTACCGCGCGCTCTACTCTGTGAGGACCGCAAGCTGAGCCGAG 372  
DB 129 AsnLeuThrIlePheThrAlaArgLeuTyrTrpPheGln---TyrProCysTyrGlnGlu 147  
QY 373 GGGCTGGCGGGTGTGCACCGCGCGGTGCAATCATGCATCATCATTCATAAGATTAT 432  
DB 148 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrGluAspPhe 167  
QY 433 TTTTACTGCTGGAATACTTTTGTAGAAACCATGAAAGAACTTCAAGACCTGGGAAGG 492  
DB 168 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly 187

```
Qy 493 CTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 188 LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu 203

RESULT 5
Q96F12 PRELIMINARY; PRT; 190 AA.
AC Q96F12;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Similar to APOBEC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RC TISSUE=UTERUS;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC011739; AAF11739.1; -.
DR InterPro: IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; Cyt_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 190 AA; 22827 MW; DA0584EF75C91CF0 CRC64;

Alignment Scores:
Pred. No.: 1,78e-28 Length: 190
Score: 367.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 33.38% Indels: 9
DB: 4 Gaps: 4

US-09-966-880A-7_COPY_80_676 (1-597) x Q96F12 (1-190)
Qy 31 TTCTTTTACCAATTCAAAATGTCCTGGCTAAGGCTGCGGTGAGACCTACCTGTGC 90
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17 PheTyrPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAspGluThrTrpLeuCys 36
Qy 91 TAGCTAGTGAAG-----AGGCGTGACAGTGCTACATCCTTTTCACCTGGACTTGGTTAT 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37 PheThrValGluGlyIleLysArgArgSerValSerTrpLysThr-----GlyVal 54
Qy 145 CTTTCGCAAT-----AAGAACGGCTGCACGTGGGAATTGCTCTTCCTCCGCTAC 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74
Qy 193 ATCTCGGACTGGACCTAGACCTGCGCGTGTCTACCGGTACCTGGTTACCTCCTGG 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTyrThrSerTrp 94
Qy 253 AGCCCTGTACAGCTGTGCCGACATGTGGCGGACTTCTGCGGAGGAGACCCCAACCTC 312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 114
Qy 313 AGTCTGAGGATCTTACCGCGCGCTACTTCTGTGAGGACCGCAGGCTGAGCCGAG 372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu 133
Qy 373 GGGCTCGGGCGGTGCACCGCGCGGTGCAAAATAGCCATCATGACCTTCAAAGATTAT 432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrGluAspPhe 153
Qy 433 TTTTACTGCTGGAATCTTTGTAGAAAACCATGAAGAAGCTTTCAAAGCTGGGAAGG 492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly 173
Qy 493 CTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu 189

RESULT 6
Q9NRW3 PRELIMINARY; PRT; 294 AA.
AC Q9NRW3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Phorbollin I protein.
GN PBI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gu J., Huang Q., Yu Y., Xu S., Wang Y., Han Z., Chen Z., Zhou J.,
RA Tu Y., Gu W., Fu G., Huang C.;
RT "Novel genes expressed in hematopoietic stem/progenitor cells from
RT Myelodysplastic Syndromes patient.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF165520; AAF86650.1; -.
DR InterPro: IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; Cyt_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 294 AA; 33363 MW; 1B39C7A13D690901 CRC64;

Alignment Scores:
Pred. No.: 4,44e-27 Length: 294
Score: 354.00 Matches: 75
Percent Similarity: 59.06% Conservative: 26
Best Local Similarity: 43.86% Mismatches: 60
Query Match: 32.15% Indels: 10
DB: 4 Gaps: 5

US-09-966-880A-7_COPY_80_676 (1-597) x Q9NRW3 (1-294)
Qy 31 TTCTTTTACCAATTCAAAATGTCCTGGCTAAGGCTGCGGTGAGACCTACCTGTGC 90
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 17 PheTyrPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys 36
Qy 91 TAGCTAGTGAAG-----AGGCGTGACAGTGCTACATCCTTTTCACCTGGACTTGGTTAT 144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 37 PheThrValGluGlyIleLysArgArgSerValSerTrpLysThr-----GlyVal 54
Qy 145 CTTTCGCAAT-----AAGAACGGCTGCACGTGGGAATTGCTCTTCCTCCGCTAC 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74
Qy 193 ATCTCGGACTGGACCTAGACCTGCGCGTGTCTACCGGTACCTGGTTACCTCCTGG 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 75 PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTyrThrSerTrp 94
Qy 253 AGCCCTGTACAGCTGTGCCGACATGTGGCGGACTTCTGCGGAGGAGACCCCAACCTC 312
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 114
Qy 313 AGTCTGAGGATCTTACCGCGCGCTACTTCTGTGAGGACCGCAGGCTGAGCCGAG 372
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 115 AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu 133
Qy 373 GGGCTCGGGCGGTGCACCGCGCGGTGCAAAATAGCCATCATGACCTTCAAAGATTAT 432
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 134 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyrGluAspPhe 153
Qy 433 TTTTACTGCTGGAATCTTTGTAGAAAACCATGAAGAAGCTTTCAAAGCTGGGAAGG 492
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly 173
Qy 493 CTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 174 Ile---LysAsnGlnLeuSerThrSerGluLys 183

RESULT 7
Q99J72
```





RA Peck A.;  
RT "APOBEC-2, a cardiac and skeletal muscle specific member of the  
RT cytidine deaminase supergene family.";  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=HEART;  
RX MEDLINE=99333690; PubMed=10403781;  
RA Liao W., Hong S.-H., Chang B.-J., Rudolph F.B., Clark S.C., Chan L.;  
RT "APOBEC-2, a cardiac- and skeletal muscle-specific member of the  
RT cytidine deaminase supergene family.";  
RL Biochem. Biophys. Res. Commun. 260:398-404(1999).  
DR EMBL; AL031778; CAB4740.1; -;  
DR EMBL; AF161698; AAD45360.1; -;  
KW Lipoprotein.  
SQ SEQUENCE 224 AA; 25703 MW; CA0905AF8A8C8FA1 CRC64;

Alignment Scores:  
Pred. No.: 4,96e-19 Length: 224  
Score: 274.00 Matches: 64  
Percent Similarity: 50.28% Conservative: 25  
Best Local Similarity: 36.16% Mismatches: 76  
Query Match: 24.89% Indels: 12  
DB: 4 Gaps: 5

US-09-966-880A-7\_COPY\_80\_676 (1-597) x Q9Y235 (1-224)

QY 31 TTTCTTTACCAATTCAAAATGTCGCTGGGTGAGGTGCGCGCTGAGACCTACCTGTGC 90  
Db 52 PheLysPheGlnPheArgAsnValGluTyrSerGlyArgAsnLysThrPheLeuCys 71  
QY 91 TACGTAGTGAAGAGCGGTGACAGTGCTACATCTTTTCACTGGACTTTGGTTATCTTCGC 150  
Db 72 TyrValValGluAlaGlnGlyLysGlyGlyGlnValGlnAlaSerArgGlyTyrLeuGlu 91  
QY 151 AATAAGAAC---GGCTGCCACGTGGAATGCTCTCCCTCGCTACATCTCGGACTGGGAC 207  
Db 92 AspGluHisAlaAlaHisAlaGluGluAlaPhePheAsnThrIleLeuPro---Ala 110  
QY 208 CTAGACCTGCGCGTGTACCGCTCACCTGCTCACCTCTCGGAGCCCTGCTACGAC 267  
Db 111 PheAspProAlaLeuArgTyrAsnValThrTyrValSerSerProCysAlaAla 130  
QY 268 TGTGCGCGACATGTGGCGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTC 327  
Db 131 CysAlaAspArgIleIleLysThrLeuSerLysThrLysAsnLeuArgLeuIleLeu 150  
QY 328 ACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAG-----GGGCTG 378  
Db 151 ValGlyArgLeuPheMetTrpGlu-----GluProGluIleGlnAlaAlaLeu 166  
QY 379 CGGCGCTGCACCGCGCGGTGCAATAGCATCATGACCTTCAAGATTATTTTAC 438  
Db 167 LysLysLeuLysGluAlaGlyCysLysLeuArgIleMetLysProGlnAspPheGluTyr 186  
QY 439 TGCTGGAATCTTTGTAGAAACCATGAA-----AGAACTTTCAAGCCCTGGGAA 489  
Db 187 ValTrpGlnAsnPheValGluGlnGluGlyGluSerLysAlaPheGlnProTrpGlu 206  
QY 490 GGGTGCATGAAATTCAGTTCGTCTCCAGACAGCTTCGCGCATCTT 540  
Db 207 AspIleGlnGluAsnPheLeuTyrTyrGluGluLysLeuAlaAspIleLeu 223

RESULT 12  
QY553 PRELIMINARY; PRT; 121 AA.  
AC QY553;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE BK150C2.10 (Putative novel Phorbol 1 like protein) (Fragment).  
GN BK150C2.10.  
OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Matthews L.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL022318; CAB45273.1; -;  
FT NON\_TER 1 1  
FT NON\_TER 121 121  
SQ SEQUENCE 121 AA; 14309 MW; 97FC47DCDD82B247 CRC64;

US-09-966-880A-7\_COPY\_80\_676 (1-597) x Q9Y553 (1-121)

QY 181 TTCCTCGGTACATCTCGGACTCGGACTAGACCTGCGCGTCTACCGGCTCACCTGG 240  
Db 3 PheLeuSerTrpPheCysAspIleLeuSerProAsnThrAsnTyrGluValThrTrp 22  
QY 241 TTCACCTCTGGAGCCCTGCTACGACTGTGCGCGGACATGTGGCGGACTTCTCGGAGG 300  
Db 23 TyrThrSerTrpSerProCysAlaGlyGluValAlaGluPheLeuAlaArg 42  
QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
Db 43 HisSerAsnValAsnLeuThrIlePheThrAlaArgLeuCysTyrPheTrpAspThrAsp 62  
QY 361 GCTGAGCCCGAGGGCTGCGCGCTGCACCGCGCGGGTGCNAATAGCCATCATGACC 420  
Db 63 TyrGln---GluGlyLeuCysSerLeuSerGlnGlyAlaSerValLysIleMetGly 81  
QY 421 TTCAAAGATTATTTTACTGCTGTAATCTTTGTAGAAAACCATGAAGACTTCAAA 480  
Db 82 TyrLysAspPheValSerCysTrpLysAsnPheValTyrSerAspGluProPheLys 101  
QY 481 GCTGGAAGGGGTGCTGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540  
Db 102 ProTrpLysGlyLeuGlnThrAsnPheArgLeuLeuLysArgArgLeuArgGluIleLeu 121

RESULT 13

QY550 PRELIMINARY; PRT; 103 AA.  
AC QY550;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DE BK150C2.6 (Putative novel protein similar to APOBEC1 (Apolipoprotein B mRNA editing protein) and Phorbol) (Fragment).  
DE BK150C2.6.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Matthews L.;  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL022318; CAB45276.1; -;  
DR InterPro; IPR002125; dCMP/cyt\_deam.  
DR PROSITE; PS00903; CYT\_DCMP\_DEAMINASES; UNKNOWN\_1.  
KW Lipoprotein.  
FT NON\_TER 1 1  
FT NON\_TER 103 103  
SQ SEQUENCE 103 AA; 12146 MW; 5DC969AE3ED348D9 CRC64;

Alignment Scores:

```

Pred. No.: 1,99e-17 Length: 103
Score: 257.50 Matches: 49
Percent Similarity: 63.27% Conservative: 13
Best Local Similarity: 50.00% Mismatches: 35
Query Match: 23.39% Indels: 1
DB: 4 Gaps: 1

US-09-966-880a-7_COPY_80_676 (1-597) x Q9Y550 (1-103)

QY 163 TGCCAGGTGAATGCTCTCTCGCTACATCTCGGACTGGAGCTAGACCTGGCCGC 222
Db ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
7 CysHisAlaGluArgCysPheLeuSerTrpPheCysAspPheLeuSerProAsnThr 26
QY 223 TGCTACCGGCTCACCTGGTTCACCTCTGGAGCCCTGCTAGACTGTGCGGACATGG 282
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
27 AsnTyrGluValThrTrpTyrTrpSerTrpSerProCysProGluCysAlaGlyGluVal 46
QY 283 GCGGACTTCTCGGAGGAGACCCCAACCTCAGCTGAGGATCTTACCGCGCGCTCTAC 342
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
47 AlaGluPheLeuAlaArgHisSerAsnValAsnLeuThrPheThrAlaArgLeuTyr 66
QY 343 TTCTGTGAGGACCGCAAGCTGAGCCCGGAGGGCTGCGCGCTGCACCGCGCGGGTG 402
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 TyrPheTrpAspThrAspTyrGln--GluGlyLeuArgSerLeuSerGlnGluGlyAla 85
QY 403 CAATAGCATCATGACCTTCAAGATATTTTACTGCTGGAATATTTTGTGTA 456
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
86 SerValGluIleMetGlyTyrLysAspPheLysTyrCysTrpGluAsnPheVal 103

RESULT 14
Q9UH18
ID Q9UH18 PRELIMINARY: PRT: 111 AA.
AC Q9UH18;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE BK150C2.1 (Putative novel Phorbol 1 like protein) (Fragment).
GN BK150C2.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022318; CAB45269.1; -.
FT NON_TER 1
SQ SEQUENCE 111 AA; 13006 MW; EF8922ABFFA7028 CRC64;

Alignment Scores:
Pred. No.: 9.13e-17 Length: 111
Score: 251.00 Matches: 51
Percent Similarity: 64.55% Conservative: 20
Best Local Similarity: 46.36% Mismatches: 29
Query Match: 22.80% Indels: 10
DB: 4 Gaps: 3

US-09-966-880a-7_COPY_80_676 (1-597) x Q9UH18 (1-111)

QY 226 TACCGGCTCACCTGGTTCACCTCTCGGAGCCCTGCTAGCAGC-----TGTCGCGACAT 279
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2 TyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGlyGlu 21
QY 280 GTGGCGGATTTCTGCGAGGAGAACCCCAACCTCAGCTGAGGATCTTACCGCGCGCTC 339
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
22 ValArgAlaPheLeuGlnGluAnThrHisValArgLeuArgIlePheAlaAlaArgIle 41
QY 340 TACTTCTGTGAGGACCGCAAGCTGAGCC-----GAGGGCTGCGCGGCTGCGAC 390
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
42 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 56
QY 391 CGCGCGGGGTGCAATAGCCATCATGACCTTCAAGATATTTTACTGTGGAATACT 450

```

Search completed: June 14, 2003, 18:47:43

```

Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
57 AspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheLysHisCysTrpAspThr 76
QY 451 TTTGTAGAAACCATGAAAGAACTTTCAAAGCCTCGGAAGGCTGCATGAAATTCAGTT 510
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
77 PheValAspHisGlnGlyCysProPheGlnProTrpAspGlyLeuAspGluHisSerGln 96
QY 511 COTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
97 AlaLeuSerGlyArgLeuArgAlaIleLeu 106

RESULT 15
Q9Y554
ID Q9Y554 PRELIMINARY: PRT: 121 AA.
AC Q9Y554;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BK150C2.9 (Putative novel Phorbol 1 like protein) (Fragment).
GN BK150C2.9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022318; CAB45272.1; -.
FT NON_TER 1
SQ SEQUENCE 121 AA; 14714 MW; B8FE133AACBE6D59 CRC64;

Alignment Scores:
Pred. No.: 1.06e-15 Length: 121
Score: 240.50 Matches: 48
Percent Similarity: 54.92% Conservative: 19
Best Local Similarity: 39.34% Mismatches: 54
Query Match: 21.84% Indels: 1
DB: 4 Gaps: 1

US-09-966-880a-7_COPY_80_676 (1-597) x Q9Y554 (1-121)

QY 175 TTGCTCTTCTCTCGCTACATCTCGGACTGGAGCTAGACCTGGCCGCTGCTACCGGTC 234
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1 MetCysPheLeuSerTrpPheCysGlyAsnArgLeuProAlaAsnArgArgPheGlnIle 20
QY 235 ACCTGTTTCACTCTCTGAGCCCTGCTAGCTGTGCGCGACATGTGCGCGACTTTCTG 294
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
21 ThrTrpPheValSerTrpAsnProCysLeuProCysValValValThrLysPheLeu 40
QY 295 CGAGGGAACCCCAACCTCAGCTGAGGATCTTACCGCGCGCTTCTTCTGTCAGGAC 354
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
41 AlaGluHisProAsnValThrLeuThrIleSerAlaAlaArgLeuTyrTyrArgAsp 60
QY 355 CGCAAGGCTGAGCCCGGAGGGCTGCGCGGCTGCACCGCGGGGTGCAATAGCCATC 414
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
61 ArgAspTrpArg---TrpValLeuLeuArgLeuHisLysAlaGlyAlaArgValLysIle 79
QY 415 ATGACCTTCAAGATATTTTACTGCTGGAATACTTTTGTAGAAACCATGAAAGAACT 474
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
80 MetAspTyrGluAspPheAlaTyrCysTrpGluAsnPheValCysAsnGluGlyGlnPro 99
QY 475 TTCAAAGCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGC 534
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
100 PheMetProTrpTyrLysPheAspAsnTyrAlaSerLeuHisArgThrLeuLysGlu 119
QY 535 ATCCTT 540
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
120 IleLeu 121

```

Job time : 40.6063 secs

---





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 14, 2003, 18:13:03 ; Search time 5.75003 Seconds  
(without alignments)  
6109.704 Million cell updates/sec

Title: US-09-966-880A-7\_COPY\_80\_676

Perfect score: 1101  
Sequence: 1 atggacagccttgatgaa.....ttcgtacttggactttga 597

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_n2p model -DEV=xlp  
-Q/cgn2\_1/USPTO.spool/US09966880/runat\_14062003\_175525\_10370/app\_query.fasta\_1.9493  
-DB=issued\_Patents\_AA -QFFT=fastan -SUFFIX=n2p.ra -MINMATCH=0.1 -LOOPEL=0  
-DOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09966880 @cgn\_1.1.139 @runat\_14062003\_175525\_10370 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	369.5	33.6	190	1	US-08-816-241-1
2	369.5	33.6	190	3	US-09-128-395-1
3	271	24.6	222	1	US-08-687-895-1
4	271	24.6	222	2	US-09-040-482-1
5	245	22.3	116	1	US-08-687-895-3
6	245	22.3	116	2	US-08-816-241-3
7	245	22.3	116	2	US-09-040-482-3
8	245	22.3	116	3	US-09-128-395-3
9	216.5	19.7	236	1	US-08-687-895-4
10	216.5	19.7	236	1	US-08-816-241-4
11	216.5	19.7	236	2	US-09-040-482-4
12	216.5	19.7	236	3	US-09-128-395-4

13	210	19.1	236	1	US-08-158-682A-4	Sequence 4, Appli
14	198	18.0	229	1	US-08-158-682A-2	Sequence 2, Appli
15	198	18.0	229	1	US-08-015-203-2	Sequence 2, Appli
16	198	18.0	229	1	US-08-687-895-5	Sequence 5, Appli
17	198	18.0	229	1	US-08-816-241-5	Sequence 5, Appli
18	198	18.0	229	2	US-09-040-482-5	Sequence 5, Appli
19	198	18.0	229	3	US-09-128-395-5	Sequence 5, Appli
20	88.5	8.0	1513	5	PCT-US93-03076-2	Sequence 2, Appli
21	85.5	7.8	786	4	US-09-103-429A-3	Sequence 3, Appli
22	85.5	7.8	805	4	US-09-103-429A-4	Sequence 3, Appli
23	84.5	7.7	1128	4	US-09-627-650B-11	Sequence 11, Appli
24	84.5	7.7	1128	4	US-09-436-063C-11	Sequence 11, Appli
25	84.5	7.7	1652	4	US-09-627-650B-1	Sequence 1, Appli
26	84.5	7.7	1652	4	US-09-436-063C-1	Sequence 1, Appli
27	84.5	7.7	2508	4	US-09-627-650B-7	Sequence 7, Appli
28	84.5	7.7	2508	4	US-09-436-063C-7	Sequence 7, Appli
29	84.5	7.7	2544	4	US-09-627-650B-3	Sequence 3, Appli
30	84.5	7.7	2544	4	US-09-436-063C-3	Sequence 3, Appli
31	82.5	7.5	801	1	US-07-906-349A-6	Sequence 6, Appli
32	81.5	7.4	1185	4	US-09-041-886-23	Sequence 23, Appli
33	81	7.4	440	3	US-09-092-315-3	Sequence 3, Appli
34	81	7.4	575	3	US-08-922-865-2	Sequence 2, Appli
35	81	7.4	575	4	US-09-510-949-2	Sequence 2, Appli
36	81	7.4	1312	4	US-09-041-886-19	Sequence 19, Appli
37	80	7.3	267	4	US-08-818-112-142	Sequence 142, App
38	80	7.3	267	4	US-08-818-111-137	Sequence 137, App
39	80	7.3	267	4	US-09-056-556-142	Sequence 142, App
40	80	7.3	267	4	US-09-072-596-137	Sequence 137, App
41	80	7.3	476	4	US-09-092-315-5	Sequence 5, Appli
42	79.5	7.2	793	4	US-09-588-256-10	Sequence 10, Appli
43	79	7.6	135	4	US-09-419-485-1	Sequence 1, Appli
44	78.5	7.1	521	2	US-08-406-855A-19	Sequence 19, Appli
45	78.5	7.1	521	3	US-09-206-899-19	Sequence 19, Appli

ALIGNMENTS

RESULT 1  
US-08-816-241-1  
; Sequence 1, Application US/08816241  
; Patent No. 5804185  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surva K.  
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/816,241  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0239 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 1:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT09
; CLONE: 1646823
US-08-816-241-1

Alignment Scores:
Pred. No.: 3.67e-35 Length: 190
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 33.56% Indels: 9
DB: 1 Gaps: 4

US-09-966-880A-7_COPY_80_676 (1-597) x US-08-816-241-1 (1-190)
QY 31 TTCTTTTACCAATTCAAAATGTCGGCTGAGGTGCGGTGAGACCTACCTGTGC 90
Db 17 PheTyrPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys 36
QY 91 TACGTAGTGAAG-----AGCGGTGACAGTGTACATCTTTTCTACCTGGACTTGGTTAT 144
Db 37 PheThrValGluGlyIleLysArgSerValSerTrpLysThr-----GlyVal 54
QY 145 CTTCGCAAT-----AAGACGGCTGCCAGTGGAAATGCTCTCTCCGCTAC 192
Db 55 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74
QY 193 ATCTCGAGCTGGACCTAGACCTGCGCTGCTACCGCTGCTACCGGTGAGACCTACCTGTGC 252
Db 75 PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTrpLysThr 94
QY 253 ACCCTCTGTACGATGTCGCCGACATGTGGCGGACTTTCTCGAGGAGCAACCTC 312
Db 95 SerProCysProAspCysAlaGlyValAlaGluPheLeuAlaArgHisSerAsnVal 114
QY 313 AGCTGTGAGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAGCTGAGCCGAG 372
Db 115 AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu 133
QY 433 TTTTACTGCTGAATCTTTGTAGAAACCATGAAGACTTTCAAAGCTGCGGAAGG 492
Db 154 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly 173
QY 493 CTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 174 LeuLysThrAsnPheArgLeuLeuLysArgLeuArgLeuSerLeu 189

RESULT 2
US-09-128-395-1
; Sequence 1, Application US/09128395
; Patent No. 6087108
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,395
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/816,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: PROSTUT09
; CLONE: 1646823
US-09-128-395-1

Alignment Scores:
Pred. No.: 3.67e-35 Length: 190
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 33.56% Indels: 9
DB: 1 Gaps: 4

US-09-966-880A-7_COPY_80_676 (1-597) x US-09-128-395-1 (1-190)
QY 31 TTCTTTTACCAATTCAAAATGTCGGCTGAGGTGCGGTGAGACCTACCTGTGC 90
Db 17 PheTyrPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys 36
QY 91 TACGTAGTGAAG-----AGCGGTGACAGTGTACATCTTTTCTACCTGGACTTGGTTAT 144
Db 37 PheThrValGluGlyIleLysArgSerValSerTrpLysThr-----GlyVal 54
QY 145 CTTCGCAAT-----AAGACGGCTGCCAGTGGAAATGCTCTCTCCGCTAC 192
Db 55 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 74
QY 193 ATCTCGAGCTGGACCTAGACCTGCGCTGCTACCGCTGCTACCGGTGAGACCTACCTGTGC 252
Db 75 PheCysAspAspIleLeuSerProAsnThrLysTyrGlnValThrTrpTrpLysThr 94
QY 253 AGCCCTCTGTACGATGTCGCCGACATGTGGCGGACTTTCTCGAGGAGCAACCTC 312
Db 95 SerProCysProAspCysAlaGlyValAlaGluPheLeuAlaArgHisSerAsnVal 114
QY 313 AGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAGCTGAGCCGAG 372
Db 115 AsnLeuThrIlePheThrAlaArgLeuTyrTyrPheGln---TyrProCysTyrGlnGlu 133
QY 373 GGGCTGCGGCGCTGCACCGCGCGGTGCAAAATAGCATCATGACCTTCAAAGATTAT 432
Db 134 GlyLeuArgSerLeuSerGlnGluGlyValAlaValAlaGluIleMetAspTyrGluAspPhe 153
QY 433 TTTTACTGCTGAATCTTTGTAGAAACCATGAAGACTTTCAAAGCTGCGGAAGG 492
Db 154 LysTyrCysTrpGluAsnPheValTyrAsnAspAsnGluProPheLysProTrpLysGly 173
QY 493 CTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540

```



```
US-09-966-880A-7_COPY_80_676 (1-597) x US-09-040-482-1 (1-222)
QY 31 TTTCTTTTACCAATTCAAAATGTCGCTGGCGTAAAGGTCGGCGTGAGACCTACCTGTGC 90
Db 52 PheLysPheGlnPheArgAsnValGluTyrSerGlyArgAsnLysThrPheLeuCys 71
QY 91 TAGCTAGTGAAGAGCGGTGACAGTGTACATCCCTTTTTCACCTGGAGCTTTGGTTATCTTCGC 150
Db 72 TyrValValGluAlaGlnGlyLysGlyGlnValcInAlaSerArgGlyTyrLeuGlu 91
QY 151 AATAAGAAC---GGCTGCCACGTGGGAATGCTCTTCCTCGGTACATCTCGGACTGGAC 207
Db 92 AspGluHisAlaAlaHisAlaGluAlaPhePheAsnThrileLeuPro---Ala 110
QY 208 CTAGACCTCGGCGCTACCGCGTCACCTGCTTCCCTCGGTGGAGCCCTCTCAGGAC 267
Db 111 PheAspProAlaLeuArgTyrAsnValThrTrpTyrValSerSerProCysAlaAla 130
QY 268 TGTGCCGACATGTGCCGACTTTCGCGAGGAACCCCAACCTCATCTGAGGATCTTC 327
Db 131 CysAlaAspArgile**LysThrLeuSerLysThrLysAsnLeuArgLeuLeu 150
QY 328 ACCGCGGCTCTACTCTGTGAGGACCGCAGGCTGAGCCGAG-----GGCTG 378
Db 151 ValGlyArgLeuPheMetTrpGlu-----GluProGluLeGlnAlaAlaLeu 166
QY 379 CGCGCGCTGCACGCGCGCGGTGCAATAGCCATCATGACCTTCAAGATTATTTTAC 438
Db 167 LysLysLeuLysGluAlaGlyCysLysLysLeuArgIleMetLysProGlnaspPheGluTyr 186
QY 439 TGTCTGAATCTTTGTAGAAACCATGAA-----AGAACTTTCAAGCCTCGGAA 489
Db 187 ValTrpGlnAsnValGluGlnGluGluSerLysAlaPheGlnProTrpGlu 206
QY 490 GGCTGCATCAAAATCAGTTCCTCTCCAGACAGCTTCGGCGCATC 537
Db 207 AspIleGlnGluAsnPheLeuTyrTyrGluGluLysLeuAlaAspIle 222

RESULT 5
US-08-816-241-3
; Sequence 3, Application US/08687895
; Patent No. 5747319
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,895
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
```

```
; LENGTH: 116 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 436941
US-08-687-895-3

Alignment Scores:
Pred. No.: 1,4e-20 Length: 116
Score: 245.00 Matches: 51
Percent Similarity: 61.82% Conservative: 17
Best Local Similarity: 46.36% Mismatches: 32
Query Match: 22.25% Indels: 10
DB: 1 Gaps: 3

US-09-966-880A-7_COPY_80_676 (1-597) x US-08-687-895-3 (1-116)
QY 226 TACCGCGTCACCTGTGTTCACCTCTCCCTGAGCCCTGCTACGAC-----TGTGCCCGACAT 279
Db 7 TyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGlyGlu 26
QY 280 GTGGCGGACTTTCTGCGAGGGAACCCCAACCTCAGTCTCAGGATCTTCACCGCGCGCTC 339
Db 27 ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgIle 46
QY 340 TACTTCTGTGAGGACCGCAGGCTGAGCCC-----GAGGGGCTCGCGCGCTGCAC 390
Db 47 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 61
QY 391 CGCGCGCGGTGCAATAGCCATCATGACCTTCAAGATTATTTTACTGCTGGAATACT 450
Db 62 AspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheGluTyrCysTrpAspThr 81
QY 451 TTTGTAGAAACCATGAAAGAACTTTCAAGGCTGGGAAGGCTGCATGAAATTCAGTT 510
Db 82 PheValTyrArgGlnGlyCysProPheGlnProTrpAspGlyLeuGluHisSerGln 101
QY 511 CGTCTCTCAGACAGCTTCGGCGCATCCTT 540
Db 102 AlaLeuSerGlyArgLeuArgAlaIleLeu 111

RESULT 6
US-08-816-241-3
; Sequence 3, Application US/08816241
; Patent No. 5804185
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,241
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
```

NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 116 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 436941  
US-08-816-241-3

Alignment Scores:  
Pred. No.: 1.4e-20 Length: 116  
Score: 245.00 Matches: 51  
Percent Similarity: 61.82% Conservative: 17  
Best Local Similarity: 46.36% Mismatches: 32  
Query Match: 22.25% Indels: 10  
DB: 3 Gaps: 3

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-08-816-241-3 (1-116)

QY 226 TACCGCGTCACCTGGTTACCTCGAGCCCTGCTAGCAC-----TGTCGCCGACAT 279  
Db 7 TyrArgValThrPheIleSerTrpSerProcysPheSerTrpGlyCysAlaGlyGlu 26  
QY 280 GTGGCCGACCTTCTCGAGGGAAACCCCAACCTCAGTCAGGATCTTCACCGCGCCCTC 339  
Db 27 ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgIle 46  
QY 340 TACTTCTGTGAGGACCGAAGCTGAGCCC-----GAGGGCTGCGGGCGCTGCAC 390  
Db 47 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 61  
QY 391 CGCGCGGGGTGCAATAGCCATCAGCTTCAAGATTATTTTACTGCTGGAATACT 450  
Db 62 AspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheGluTyrCysTrpAspThr 81  
QY 451 TTTGTAGAAAACCATGAAAGAACTTTCAAAGCTTGGGAAGGCTGCATGAAATTCAGTT 510  
Db 82 PheValTyrArgGlnGlyCysProPheGlnProTyrAspGlyLeuGluHisSerGln 101  
QY 511 CGTCTCTCCAGACAGCTTCGGCGCATCCTT 540  
Db 102 AlaLeuSerGlyArgLeuArgAlaIleLeu 111

RESULT 7  
US-09-480-482-3  
Sequence 3, Application US/09040482  
Patent No. 5916556  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Hawkins, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/040,482  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/687,895  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0109 US  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 116 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 436941  
US-09-040-482-3

Alignment Scores:  
Pred. No.: 1.4e-20 Length: 116  
Score: 245.00 Matches: 51  
Percent Similarity: 61.82% Conservative: 17  
Best Local Similarity: 46.36% Mismatches: 32  
Query Match: 22.25% Indels: 10  
DB: 3 Gaps: 3

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-09-040-482-3 (1-116)

QY 226 TACCGCGTCACCTGGTTACCTCGAGCCCTGCTAGCAC-----TGTCGCCGACAT 279  
Db 7 TyrArgValThrPheIleSerTrpSerProcysPheSerTrpGlyCysAlaGlyGlu 26  
QY 280 GTGGCCGACCTTCTCGAGGGAAACCCCAACCTCAGTCAGGATCTTCACCGCGCCCTC 339  
Db 27 ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgIle 46  
QY 340 TACTTCTGTGAGGACCGAAGCTGAGCCC-----GAGGGCTGCGGGCGCTGCAC 390  
Db 47 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 61  
QY 391 CGCGCGGGGTGCAATAGCCATCAGCTTCAAGATTATTTTACTGCTGGAATACT 450  
Db 62 AspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheGluTyrCysTrpAspThr 81  
QY 451 TTTGTAGAAAACCATGAAAGAACTTTCAAAGCTTGGGAAGGCTGCATGAAATTCAGTT 510  
Db 82 PheValTyrArgGlnGlyCysProPheGlnProTyrAspGlyLeuGluHisSerGln 101  
QY 511 CGTCTCTCCAGACAGCTTCGGCGCATCCTT 540  
Db 102 AlaLeuSerGlyArgLeuArgAlaIleLeu 111

RESULT 8  
US-09-128-395-3  
Sequence 3, Application US/09128395  
Patent No. 6087108  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA

```

; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,395
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/816,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 116 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 436941
US-09-128-395-3

```

```

Alignment Scores:
Pred. No.: 1,4e-20 Length: 116
Score: 245.00 Matches: 51
Percent Similarity: 61.82% Conservative: 17
Best Local Similarity: 46.36% Mismatches: 32
Query Match: 22.25% Indels: 10
DB: 3 Gaps: 3

```

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-09-128-395-3 (1-116)

```

QY 226 TACGCGGTACCTGGTTCACCTCCTGGAGCCCTGCTAGCAC-----TGTGCGCGACAT 279
Db 7 TyrArgValThrTrpPheIleSerTrpSerProCysPheSerTrpGlyCysAlaGlyGlu 26
QY 280 GTGCGCGACTTTCGCGAGGAACCCCAACCTCAGTCTGAGGATCTTCACGCGCGCTC 339
Db 27 ValArgAlaPheLeuGlnGluAsnThrHisValArgLeuProIlePheAlaAlaArgIle 46
QY 340 TACTTCTCTGAGGACCGCAAGGCTGAGCCC-----GAGGGGCTGCGCGCGCTGCAC 390
Db 47 Tyr-----AspTyrAspProLeuTyrLysGluAlaLeuGlnMetLeuArg 61
QY 391 CGCGCCGGGTGCAATAGCATCATGACCTTCAAGATATATTTTACTGCTGCTGAATACT 450
Db 62 AspAlaGlyAlaGlnValSerIleMetThrTyrAspGluPheGluTyrCysTrpAspThr 81
QY 451 TTTGTAGAAAACCATGAAGAACTTCAAAAGCCTGGGAAGGCTGCATGAAATACGTT 510
Db 82 PheValTyrArgGlnGlyCysProPheGlnProTrpAspGlyLeuGluGluHisSerGln 101
QY 511 CGTCTCTCCAGACGCTTCGGCGCATCCTT 540
Db 102 AlaLeuSerGlyArgLeuArgAlaIleLeu 111

```

# RESULT 9

```

US-08-687-895-4
; Sequence 4, Application US/08687895
; Patent No. 5747319
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.

```

```

; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,895
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1177798
US-08-687-895-4

```

```

Alignment Scores:
Pred. No.: 4.29e-17 Length: 236
Score: 216.50 Matches: 44
Percent Similarity: 59.52% Conservative: 31
Best Local Similarity: 34.92% Mismatches: 42
Query Match: 19.66% Indels: 9
DB: 1 Gaps: 4

```

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-08-687-895-4 (1-236)

```

QY 103 AGCGGTACAGTGTACATCCTTTTCACTGGACTTTGGTTATCTTCGC----- 150
Db 33 ArgLysGluAlaCysLeuLeuTyrGluIleLysTrpGlyMetSerArgLysIleTrpArg 52
QY 151 -----AATAAGAACGGCTGC---CAGCTGGAATTGCTCTCCTC---CGCTACATCTCG 198
Db 53 SerSerGlyLysAsnThrThrAsnHisValGluValAsnPheIleLysLysPheThrSer 72
QY 199 GACTGGGACCTAGACCTCGCCGCTGCTACCGCTCACCTGGTTCACCTCCTCGAGCCCC 258
Db 73 GluArgAspPheHisProSerIleSerCysSerIleThrTrpPheLeuSerTrpSerPro 92
QY 259 TGCTACGACTGTCCCGACATGTGGCGGACTTTCTCGGAGGGGAACCCCAACCTCAGTCTG 318
Db 93 CysTrpGluCysSerGlnAlaIleArgGluPheLeuSerArgHisProGlyValThrLeu 112
QY 319 AGGATCTTCAACCGCGCTCTACTTCTGTGAGACCGCAAGCTGAGCCGAGGGGCTG 378
Db 113 ValIleTyrValAlaArgLeuPheTrpHisMetAspGlnGln---AsnArgGlnGlyLeu 131
QY 379 CGCGCGCTGCACCGCGCGGGTGCAAATAGCATCATGACCTTCAAGATATTTTAC 438
Db 132 ArgAspLeuValAsnSerGlyValThrIleGlnIleMetArgAlaSerGluTyrTyrHis 151
QY 439 TGCTGGAATACTTTTCTTA 456

```

Db 152 CysTrpArgAsnPheVal 157

RESULT 10  
US-08-816-241-4  
; Sequence 4, Application US/08816241  
; Patent No. 5804195  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/816,241  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0239 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 236 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 11777906  
US-08-816-241-4

Alignment Scores:

Pred. No.:	Length:
Score: 236	Matches: 236
Percent Similarity: 59.52%	Conservative: 31
Best Local Similarity: 34.92%	Mismatches: 42
Query Match: 19.66%	Indels: 9
DB:	Gaps: 1

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-08-816-241-4 (1-236)

Qy 103 AGCGGTGACAGTCTCATCTTTTCATCGACTTGTTGGTTATCTTCGC----- 150  
|||:::|||||::: |||  
Db 33 ArgLysGluAlaCysLeuLeuTyrgluileLysTrpGlyMetSerArgLysIleTrpArg 52

Qy 151 -----ATAAGAACGGTGC---CACGTGGAATTGCTCTTCCTC---CGTACATCTCG 198  
||||| |||:::|||||::: |||  
Db 53 SerSerGlyLysAsnThrThraSnHlsValGluValAsnPheIleLysLysPheThrSer 72

Qy 199 GACTGGGACCTAGACCCTGGCGCGTGTCTACGCCGTGCCTCCCTCGTTCACCTCTCGGAGCCCC 258  
::: ||| |||  
Db 73 GluArgAspPheHisProSerIleSerCysSerIleThrTrpPheLeuSerTrpSerPro 92

Qy 259 TGCTAGCACTGTGCCCGACATGTGCGCAGCTTTCTGCGAGGAGACCCCAACCTCAGTCTG 318  
|||||:::|||||::: |||  
Db 93 CysTrpGluCysSerGlnAlaIleargGluPheLeuSerArgHisProGlyValThrLeu 112

Qy 319 AGGATCTTCAACCGCGCTCTACTTCTGTGAGGACCGCAAGCTGAGCCCAGAGGGGCTG 379  
|||::: |||:::|||||::: |||  
Db 113 ValIleTyrrValAlaargLeuPheTrpHisMetAspGlnGln---AsnArgGlnGlyLeu 131

Qy 379 CGCGCGCTGCACCGCGCGCGGGTGCAAAATAGCATCATCATACCTTCAAAGATTATTTTTAC 438  
||| ||| ::|::| ||| |||  
Db 132 ArgAspLeuValAsnSerGlyValThrIleGlnIleMetArgAlaSerGluTyrrHis 151

Qy 439 TGCTGGAATCTTTTCTGA 456  
||||| |||  
Db 152 CysTrpArgAsnPheVal 157

RESULT 11  
US-09-040-482-4  
; Sequence 4, Application US/09040482  
; Patent No. 5916556  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Hawkins, Phillip R.  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq version 1.5  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/040,482  
; FILING DATE:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/687,895  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0109 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 236 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; IMMEDIATE SOURCE:  
; LIBRARY: GenBank  
; CLONE: 1177798  
US-09-040-482-4

Alignment Scores:

Pred. No.:	Length:
Score: 236	Matches: 236
Percent Similarity: 59.52%	Conservative: 31
Best Local Similarity: 34.92%	Mismatches: 42
Query Match: 19.66%	Indels: 9
DB:	Gaps: 2

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-09-040-482-4 (1-236)

Qy 103 AGCGGTGACAGTCTCATCTTTTCATCGACTTGTTGGTTATCTTCGC----- 150  
|||:::|||||::: |||  
Db 33 ArgLysGluAlaCysLeuLeuTyrgluileLysTrpGlyMetSerArgLysIleTrpArg 52

Qy 151 -----ATAAGAACGGTGC---CACGTGGAATTGCTCTTCCTC---CGTACATCTCG 198  
||||| |||:::|||||::: |||  
Db 53 SerSerGlyLysAsnThrThraSnHlsValGluValAsnPheIleLysLysPheThrSer 72

Qy 199 GACTGGGACCTAGACCCTGGCGCGTGTCTACGCCGTGCCTCCCTCGTTCACCTCTCGGAGCCCC 258  
::: ||| |||  
Db 73 GluArgAspPheHisProSerIleSerCysSerIleThrTrpPheLeuSerTrpSerPro 92

Qy 259 TGCTAGCACTGTGCCCGACATGTGCGCAGCTTTCTGCGAGGAGACCCCAACCTCAGTCTG 318  
|||||:::|||||::: |||  
Db 93 CysTrpGluCysSerGlnAlaIleargGluPheLeuSerArgHisProGlyValThrLeu 112

151	QY	-----AATAAGAACGGCTGC---CACGTGGAATTGCTCTTCTC---CCGTACATCTCG	198
53	Db	SerSerGlyLysAsnThrThrAsnHisValGluValAsnPheIleLysLysPheThrSer	72
199	QY	GACTGGGACTAGACCGCTGCGCGTGTACCGCTACCTGGTTCACCTCCCTGGAGCCCC	258
		:::	
73	Db	GluArgAspPheHisProSerIleSerCysSerIleThrTrpPheLeuSerTrpSerPro	92
		:::	
259	QY	TGCTACGACTGTGCCGACATGTGCGCGACTTCTGCGAGGGAACCCCAACCTCAGCTCG	318
		:::	
93	Db	CysTrpGluCysSerGlnAlaIleArgGluPheLeuSerArgHisProGlyValThrLeu	112
		:::	
319	QY	AGCATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGCGTGAGCCCGAGGGCGTG	378
113	Db	ValIleTyrValAlaArgLeuPheTrpHisMetAspGlnGln---AsnArgGlnGlyLeu	131
379	QY	CGCGCGCTGCACCGCGCGGGTGCACATAGCCATGACCATGACCTTCAAGATATTTTATC	438
132	Db	ArgAspLeuValAsnSerGlyValThrIleGlnIleMetArgAlaSerGluTyrTyrHis	151
439	QY	TGCTGGAATCTTTTGTA	456
152	Db	CysTrpArgAsnPheVal	157

## RESULT 12

RESULTS 12  
US-09-128-395-4  
; Sequence 4, Application US/09128395

Sequence of Affected  
; Patent No. 6087108  
; GENERAL INFORMATION:

APPLICANT: Bandman, Olga  
APPLICANT: Goli, Surya K.

; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
 ; NUMBER OF SEQUENCES: 5

NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174  
CITY: Palo Alto

STATE: CA  
COUNTRY: USA

```

;
; ZIP: 94304
;
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible

```

```

; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ for Windows Ver

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/128,

```

FILING DATE:  
CLASSIFICATION:

; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/816,

;  
; FILING DATE:  
;  
; ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36

REFERENCE/DOCKET I  
TELECOMMUNICATION I

TELEPHONE: \_\_\_\_\_  
TELEFAX: \_\_\_\_\_

Alignment Scores:

alignment scores.	
Pred. No.:	4.29e-17
Length:	236





COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/015,203  
FILING DATE: 19930209  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Cooley, Ronald B.  
REGISTRATION NUMBER: 27,187  
REFERENCE/DOCKET NUMBER: ARCD:069  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 744-0090  
TELEFAX: (312) 245-4961  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 229 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-015-203-2

Alignment Scores:  
Pred. No.: 6,42e-15 Length: 229  
Score: 198.00 Matches: 56  
Percent Similarity: 47.06% Conservative: 24  
Best Local Similarity: 32.94% Mismatches: 62  
Query Match: 17.98% Indels: 28  
DB: 1 Gaps: 6

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-08-015-203-2 (1-229)

```
QY 70 CGCGGTGAGACCTACTGTGCTAGCTAGTGAAG-----AGG 105
Db 33 ArgLysGluThrCysLeuLeuTyrGluLeuAsnTrpGlyArgHisSerIleTrpArg 52
QY 106 CGTGACAGTGCTACATCTTTTCACTGACCTTTGGTTATCTCGCAATAAGAACGGCTGC 165
Db 53 HisThrSerGlnAsnThr-----AsnLys----- 60
QY 166 CAGTGGGAATGCTCTTCCTC---CGTACATCTCGGACCTGGGACCTAGACCCCTGGCGC 222
Db 61 HisValGluValAsnPheIleGluLysPheThrThrGluArgTyrPheCysProAsnThr 80
QY 223 TGCTACCGCGTCACCTGGTTCACCTCCTGGAGCCCTGCTACGACTGTGCCCGACATGTG 282
Db 81 ArgCysSerIleThrTrpPheLeuSerTrpSerProCysGlyGluCysSerArgAlaIle 100
QY 283 CGCGACTTCTCGCGGGAGACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTAC 342
Db 101 ThrGluPheLeuSerArgTyrProHisValThrLeuPheIleTyrIleAlaArgLeuTyr 120
QY 343 TTCTGTGAGGACCGCAGGCTGAGCCCGGCGGTGGCGGTGGTGCACCGCGCGGGGTG 402
Db 121 HisHisAlaAspProArg---AsnArgGlnGlyLeuArgAspLeuIleSerSerGlyVal 139
QY 403 CAATAGCCATCATGACCTTCAAGATTATTTTACTGCTGGAACTACTTTGTAGAAAAC 462
Db 140 ThrIleGlnIleMetThrGluGlnGluSerGlyTyrCysTrpArgAsnPheValAsnTyr 159
QY 463 CATGAAAGACTTTCAAGCTTGGGAGGCTGCATGAAATTCAGTTCTGCTCTCCAGA 522
Db 160 SerProSerAsnGluAlaHisTrpProArgTyrProHisLeuTrpValArgLeu----- 177
QY 523 CAGCTTCGGCGCATCTTTTGGCCCTGTAT 552
Db 178 -----TyrValLeuGluLeuTyr 183
```

Search completed: June 14, 2003, 18:59:22

Job time : 8.75003 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 14, 2003, 18:26:00 ; Search time 19.2225 Seconds  
(without alignments)  
6641.044 Million cell updates/sec

Title: US-09-966-880A-7\_COPY\_80\_676

Perfect score: 1101  
Sequence: 1 atggacagccttgatgaa.....ttcgtactttggagcttga 597

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 817286

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+n2p.model -DEV=xlp  
-Q=/cgn2\_1/USPTO.spool/US09966880/runat\_14062003\_175526\_10413/app\_query.fasta.1.9493  
-DB=PublishedApplications\_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09966880.ecgn.1.1.105 @runat\_14062003\_175526\_10413  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100  
-LONGLONG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications\_AA:

- 1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	98.6	198	9	US-09-966-880A-8
2	1008	91.6	198	9	US-09-966-880A-2
3	390	35.4	384	10	US-09-729-674-174
4	369.5	33.6	222	10	US-09-925-300-1639

5	211	19.2	229	9	US-09-966-880A-36
6	102.5	9.3	18636	9	US-10-073-912-17
7	99.5	9.0	902	9	US-10-184-644-303
8	99.5	9.0	902	9	US-10-184-634-303
9	98.5	8.9	2558	9	US-10-184-644-103
10	98.5	8.9	2558	9	US-10-184-634-103
11	97.5	8.9	1572	9	US-10-184-644-65
12	97.5	8.9	1572	9	US-10-184-634-65
13	97	8.8	2103	9	US-10-184-644-319
14	97	8.8	2103	9	US-10-184-634-319
15	95.5	8.7	1819	9	US-10-184-644-39
16	95.5	8.7	1819	9	US-10-184-634-39
17	95	8.6	1788	9	US-10-184-644-405
18	95	8.6	1788	9	US-10-184-634-405
19	93.5	8.5	4640	9	US-10-184-644-75
20	93.5	8.5	4640	9	US-10-184-634-75
21	92.5	8.4	4040	9	US-10-123-155-425
22	92	8.4	2570	9	US-10-123-155-539
23	92	8.4	3150	9	US-10-184-644-81
24	92	8.4	3150	9	US-10-184-634-81
25	91.5	8.3	3877	9	US-10-184-644-263
26	91.5	8.3	3877	9	US-10-184-634-263
27	91	8.3	1422	9	US-10-123-155-117
28	91	8.3	1656	9	US-10-123-155-239
29	91	8.3	2163	9	US-10-123-155-127
30	90.5	8.2	3038	9	US-10-184-644-261
31	90.5	8.2	3038	9	US-10-184-634-261
32	90.5	8.2	3265	9	US-10-184-644-257
33	90.5	8.2	3265	9	US-10-184-634-257
34	90.5	8.2	3288	9	US-10-184-644-477
35	90.5	8.2	3288	9	US-10-184-634-477
36	90	8.7	591	9	US-10-060-036-165
37	90	8.2	2294	9	US-10-184-644-283
38	90	8.2	2294	9	US-10-184-634-283
39	90	8.2	3732	9	US-10-123-155-71
40	89.5	8.1	1395	9	US-10-123-155-3
41	89.5	8.1	2933	9	US-10-123-155-345
42	89	8.1	1257	9	US-10-184-644-365
43	89	8.1	1257	9	US-10-184-634-365
44	89	8.1	2362	9	US-10-184-644-15
45	89	8.1	2362	9	US-10-184-634-15

ALIGNMENTS

RESULT 1  
US-09-966-880A-8  
; Sequence 8, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966.880A  
; CURRENT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 198  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-966-880A-8

Alignment Scores:

Pred. No.: 1,82e-102 Length: 198  
Score: 1086.00 Matches: 198  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 98.64% Indels: 0  
DB: 9 Gaps: 0

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-09-966-880A-8 (1-198)

```
QY 1 ATGACAGCGCTTGTATGACCGAGGAGAGTTTCTTTACCAATTCAAAATGTCGCTGG 60
Db 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
QY 61 GCTAAGGCTCGGCTGAGACCTACCTGTCTAGTAGTGAAGAGCGGTGACAGTGTCTACA 120
Db 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
QY 121 TCTTTTCACTGACTTTGGTTATCTTCGCAATAAGACGGGTGCCAGTGGAAATGTCTC 180
Db 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
QY 181 TTCTCCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGCTACCTGG 240
Db 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
QY 241 TTCACCTCTCGAGCGGCTGCTACGACTGTGCGCGCTGCTACCGCTGCTGCGAGGG 300
Db 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
QY 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGGCCCTCTACTTGTGAGGACCGCAAG 360
Db 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
QY 361 GCTGAGCCCGAGGGCTGCGGCTGCACCGCGGGGTGCAAAATAGCCATCATGACC 420
Db 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
QY 421 TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAACCATGAAAGAACTTTCAAA 480
Db 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
QY 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
QY 541 TTGCCCTGTATGAGTTGATGACTTACGACAGCCATTTCTGTTGGGACTT 594
Db 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
```

## RESULT 2

US-09-966-880A-2  
; Sequence 2, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:

; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966, 880A  
; CURRENT FILING DATE: 2001-09-28  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 198  
; TYPE: PRT

; ORGANISM: Mus musculus  
US-09-966-880A-2

## Alignment Scores:

Pred. No.: 1,7e-94 Length: 198  
Score: 1008.00 Matches: 183  
Percent Similarity: 95.94% Conservative: 6  
Best Local Similarity: 92.89% Mismatches: 8  
Query Match: 91.55% Indels: 0  
DB: 9 Gaps: 0

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-09-966-880A-2 (1-198)

```
QY 1 ATGACAGCGCTTGTATGACCGAGGAGAGTTTCTTTACCAATTCAAAATGTCGCTGG 60
Db 1 MetAspSerLeuLeuMetLysGlnLysPheLeuTyrHisPheLysAsnValArgTrp 20
QY 61 GCTAAGGCTCGGCTGAGACCTACCTGTGTCTAGTAGTGAAGAGCGGTGACAGTGTCTACA 120
Db 21 AlaLysGlyArgHisGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
QY 121 TCTTTTCACTGAGCTTTGGTTATCTTCGCAATAAGACGGGTGCCAGTGGAAATGTCTC 180
Db 41 SerCysSerLeuAspPheGlyHisLeuArgAsnLysSerGlyCysHisValGluLeuLeu 60
QY 181 TTCTCCCGCTACATCTCGGACTGGGACCTAGACCTGGCCGCTGCTACCGCGCTCACCTGG 240
Db 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
QY 241 TTCACCTCTCGAGCGGCTGCTACGACTGTGCGCGACATGTGCGCGACTTTCTCGGAGGG 300
Db 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaGluPheLeuArgTrp 100
QY 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGGCCCTCTACTTGTGAGGACCGCAAG 360
Db 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
QY 361 GCTGAGCCCGAGGGCTGCGGCTGCACCGCGCGGGGTGCAAAATAGCCATCATGACC 420
Db 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleGlyIleMetThr 140
QY 421 TTCAAAGATTATTTTACTGCTGGAATCTTTGTAGAAACCATGAAAGAACTTTCAAA 480
Db 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnArgGluArgThrPheLys 160
QY 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 540
Db 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuThrArgGlnLeuArgArgIleLeu 180
QY 541 TTGCCCTGTATGAGTTGATGACTTACGACAGCGCATTTCTGTTGGGA 591
Db 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgMetLeuGly 197
```

## RESULT 3

US-09-729-674-174  
; Sequence 174, Application US/09729674  
; Patent No. US20010039335A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallie, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steining, Robert J.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fectel, Kim  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM

FILE REFERENCE: 6055-64X  
CURRENT APPLICATION NUMBER: US/09/729,674  
PRIOR FILING DATE: 2000-12-04  
PRIOR APPLICATION NUMBER: 09/539,330  
PRIOR FILING DATE: 2000-03-30  
NUMBER OF SEQ ID NOS: 283  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 174  
LENGTH: 384  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-729-674-174

Alignment Scores:  
Pred. No.: 3,02e-31 Length: 384  
Score: 390.00 Matches: 83  
Percent Similarity: 61.62% Conservative: 31  
Best Local Similarity: 44.86% Mismatches: 59  
Query Match: 35.42% Indels: 12  
DB: 10 Gaps: 3

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-09-729-674-174 (1-384)

QY 16 ATGACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCGGCTGAGGCTCGCGGT 75  
DB 197 MetAspProThrPheThrPheAsnPheAsnAsnGluProTrpValArgGlyArgHis 216  
QY 76 GAGACCTACCTGCTACCTGACGTGAGAGCGGTGACAGTCTACATCCCTTTTCACCTGGAC 135  
DB 217 GluThrTyLeuCysTyGluValGluArgMethHisAsnAspThrTrpValLeuLeuAsn 236  
QY 136 -----TTTGGTTATCTTCGCAATAAG-----AACGGCTGC 165  
DB 237 GlnArgArgGlyPheLeuLeuCysAsnGlnAlaProHisLysHisGlyPheLeuGluGlyArg 256  
QY 166 CAGGTGGAATGCTCTTCTCCGCTACATCTCGGACTGGGACCTAGACCTGCGCGCTGC 225  
DB 257 HisAlaGluLeuCysPheLeuAspValIleProPheTrpLysLeuAspGlnAsp 276  
QY 226 TACCGGCTCACCTGGTTACCTCTGAGCCCTGCTAGACGTGTCGCCGACATGCGCC 285  
DB 277 TyrArgValThrCysPheThrSerTrpSerProCysPheSerCysAlaGlnGluMetAla 296  
QY 286 GACTTCTCGAGGGAACCCACCTCAGTCTGAGGATCTTACCGCGGCGCTCTACTTC 345  
DB 297 LysPheIleSerLysAsnLysHisValSerLeuCysIlePheThrAlaGlyIleTyr--- 315  
QY 346 TGTGAGGACCGCAAGCTGAGCCCGAGGGGCTCGCGGCTGACCGCGCGCGGGTGCAA 405  
DB 316 ---AspAspGlnGlyArgCysGlnGluGlyLeuArgThrLeuAlaGluAlaGlyAlaLys 334  
QY 406 ATAGCCATCATGACCTTCAAAAGATTATTTTACTGCTGGAATCTTTGTAGAAACCAT 465  
DB 335 IleSerIleMetThrTySerGluPheLysHisCysTrpAspThrPheValAspHisGln 354  
QY 466 GAAAGAACTTCAAGACCTGGGAAGGCTGCATGAAATTCAGTTGCTCTCCAGACAG 525  
DB 355 GlyCysProPheGlnProTrpAspGlyLeuAspGluHisSerGlnAspLeuSerGlyArg 374  
QY 526 CTTCCGGCGCATCCTT 540  
DB 375 LeuArgAlaIleLeu 379

RESULT 4  
US-09-925-300-1639  
Sequence 1639, Application US/09925300  
Patent No. US20020151681A1  
GENERAL INFORMATION:  
APPLICANT: Craig Rosen,  
APPLICANT: Steve Ruben  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA101  
CURRENT APPLICATION NUMBER: US/09/925,300

CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05988  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1890  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 1639  
LENGTH: 222  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-925-300-1639

Alignment Scores:  
Pred. No.: 3.18e-29 Length: 222  
Score: 369.50 Matches: 79  
Percent Similarity: 58.52% Conservative: 24  
Best Local Similarity: 44.89% Mismatches: 64  
Query Match: 33.56% Indels: 9  
DB: 10 Gaps: 4

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-09-925-300-1639 (1-222)

QY 31 TTTCTTTACCAATTCAAAAATGTCGGCTGAGGCTGAGACCTACCTGTCG 90  
DB 49 PheTyPheGlnPheLysAsnLeuTrpGluAlaAsnAspArgAsnGluThrTrpLeuCys 68  
QY 91 TAGCTAGTGAAG-----AGCGGTGACAGTCTACATCCCTTTTCACCTGGACTTTGGTTAT 144  
DB 69 PheThrValGluGlyIleLysArgSerValValSerTrpLysThr-----GlyVal 86  
QY 145 CTTTCGCAAT-----AAGACGGCTGCCACGTGGAAATGCTCTTCTCCGCTAC 192  
DB 87 PheArgAsnGlnValAspSerGluThrHisCysHisAlaGluArgCysPheLeuSerTrp 106  
QY 193 ATCTCGGACTGGGACCTAGACCTGACCGCTGCGCGCTGCTACCGCTGCTGCTGCTGCTG 252  
DB 107 PheCysAspAspIleLeuSerProAsnThrLysTyGlnValThrTrpTrpTrpSerTrp 126  
QY 253 AGCCCTGCTACGACTGTGCCGACATGTGCCGACTTCTGCGAGGGAACCCCAACCTC 312  
DB 127 SerProCysProAspCysAlaGlyGluValAlaGluPheLeuAlaArgHisSerAsnVal 146  
QY 313 AGTCTGAGGATCTTACCGCGCGCTCTACTTCTGTGAGGACCGCAAGCTGAGCCCGAG 372  
DB 147 AsnLeuThrIlePheThrAlaArgLeuTyTrpPheGln---TyrProCysTyGlnGlu 165  
QY 373 GGGCTGCGCGGCTGACCGCGCGGCTGCAATAGCCATCATGACCTTCAAGATTAT 432  
DB 166 GlyLeuArgSerLeuSerGlnGluGlyValAlaValGluIleMetAspTyGluAspPhe 185  
QY 433 TTTTACTGCTGGAATCTTTGTAGAAACCATCAAGAACTTTTCAAGCCTGGGAAGG 492  
DB 186 LysTyCysTrpGluAsnPheValTyAsnAspAsnGluProPheLysProTrpLysGly 205  
QY 493 CTCGATGAAATTCAGTCTGCTCTCCAGACAGCTTCCGCGCATCCTT 540  
DB 206 LeuLysThrAsnPheArgLeuLeuLysArgArgLeuArgGluSerLeu 221

RESULT 5  
US-09-966-880A-36  
Sequence 36, Application US/09966880A  
Patent No. US20020164743A1  
GENERAL INFORMATION:  
APPLICANT: Honjo, Tasuku  
APPLICANT: Muramatsu, Masamichi  
TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
FILE REFERENCE: 06501-088001  
CURRENT APPLICATION NUMBER: US/09/966,880A  
CURRENT FILING DATE: 2001-09-28  
PRIOR APPLICATION NUMBER: PCT/JP00/01918  
PRIOR FILING DATE: 2000-03-28  
CURRENT APPLICATION NUMBER: JP 11-371382

```

; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 36
; LENGTH: 229
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-966-880A-36

```

```

Alignment Scores:
Pred. No.: 5,07e-13 Length: 229
Score: 211.00 Matches: 49
Percent Similarity: 55.30% Conservative: 24
Best Local Similarity: 37.12% Mismatches: 49
Query Match: 19.16% Indels: 10
DB: 9 Gaps: 4

```

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-09-966-880A-36 (1-229)

```

QY 70 CGCGTGAGACCTACCTGTGCTACCTAGTGAAG-----AGCGTGACAGTGCTACATCC 123
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 33 ArgLysGluThrCysLeuLeuTyrgluileasnrtpglyGlyArgHisSerVal----- 50

QY 124 TTTTCACGTGGACTTGGTTATCTTCGCAATAAGAACGCTGCCAGTGGAAATGTCTCTC 183
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 51 -----TtpArgHisThrSerGlnAsnThrSerAsnHisValGluValAsnPhe 66

QY 184 CTC-----CGTCATCTCGGACTGGGACCTAGACCTGCGCCCTGCTACCGCGGTGCACCTGG 240
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 67 LeuGluLysPheThrGluArgTyrPheArgProAsnThrArgCysSerIleThrTrp 86

QY 241 TTCACCTCCTGGAGCCCTGTAGACTGTGCCCGACATGTGCCGACTTTCTCGGAGG 300
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 87 PheLeuSerTrpSerProCysGlyGluCysSerArgAlaIleThrGluPheLeuSerArg 106

QY 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 107 HisProTyrValThrLeuPheIleTyrIleAlaArgLeuTyrHisThrAspGlnArg 126

QY 361 GCTGAGCCCGAGGGCTGCGCGCTGCACCGCGCGGTGCAATAGCCATCATGACC 420
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 127 ---AsnArgGlnGlyLeuArgAspLeuIleSerSerGlyValThrIleGlnIleMetThr 145

QY 421 TTCAAGATATTATTTTACTGCTGGAATACTTTTGTGA 456
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 146 GluGlnGlyTyrCysTyrCysTrpArgAsnPheVal 157

```

## RESULT 6

```

US-10-073-912-17
; Sequence 17, Application US/10073912
; Publication No. US2003007703A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ34C1
; CURRENT APPLICATION NUMBER: US/10/073, 912
; CURRENT FILING DATE: 2002-02-14
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 18636
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-073-912-17

```

```

Alignment Scores:
Pred. No.: 0.239 Length: 18636
Score: 102.50 Matches: 39

```

```

Percent Similarity: 35.62% Conservative: 13
Best Local Similarity: 26.71% Mismatches: 55
Query Match: 9.31% Indels: 39
DB: 9 Gaps: 4

```

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-10-073-912-17 (1-18636)

```

QY 20 ACCGGAGGAAGTTTCTTTTACCAATTCA---AAAATGTCGCTGGCTAAGGCTCGCGTGG 76
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 14185 ThrGlyGlyAlaThrThrThrThrThrCysThrAlaAlaAlaGlyThrGlyAlaGlyThr 14204

QY 77 AGACCTACCTGTGCTACGTAGTGAAGAGCGCTGACAGTCTACATCCTTTTTCACCTGGACT 136
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 14205 ThrGlyAlaCysAlaAla----- 14210

QY 137 TTGGTTATCTTTCGCAATAAGAACGCTGCCACGTGGAATTGCTCTTCCTCGCTACATCT 196
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 14211 -----AlaCysAlaThrCysAlaCysThrGlyThrAlaAlaCys 14223

QY 197 CGGACTGGGACCTAGACCCCTGCGCTCTACCGGTACCGGTTCACCTCCCTGGA--- 253
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 14224 ThrThrGlyAlaThrThrGlyAlaCysAlaCysThrThrThrGlyThrCysAlaGlyCys 14243

QY 254 -----GCCCTGCTACGACTGTGCCGACATGTGG 283
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 14244 CysAlaAlaThrGlyThrAlaGlyThrCysAlaCysAlaThrThrAlaAlaGlyAla 14263

QY 284 CCGACTTCTCGGAGGAAACCCCACTAGTCTGAGGATCTTCACCGCGCGCTCTACT 343
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 14264 AlaThrAlaCysCysThrThrGlyThrAlaThrThrGlyAlaAlaAlaThrAlaThr 14283

QY 344 -----TCTGTGAGGACCGCAAGGCTGAGCCGAGGGCTGCGGGCGGC 385
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 14284 GlyCysThrCysCysCysThrThrThrThrThrThrThrThrThrThrThrThrThrThr 14303

QY 386 TGCACCGCGCGGGGTGC 403
    |||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db 14304 CysThrThrThrThrGlyCys 14309

```

## RESULT 7

```

US-10-184-644-303
; Sequence 303, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 303
; LENGTH: 902
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-303

```

```

Alignment Scores:
Pred. No.: 0.192 Length: 902
Score: 99.50 Matches: 41
Percent Similarity: 37.69% Conservative: 8
Best Local Similarity: 31.54% Mismatches: 68

```

```

|||||
432 GlyGlyGlyCysCysAlaGlyGlyCysAlaGlyThrGlyGlyCysAlaThr 451
|||||

Qy 83 ACCGTGCTAGCTAGTGAAGAGGGCGTGACACGTGCTACATCCTTTTCACTGGACCTTGGTT 142
|||||
452 ThrCysAlaThrGlyGlyAlaGlyAlaThrThrCys-----ThrCysCys 466
|||||

Qy 143 ATCTTCGCAATAAGACGGCTGCCACGTGGGAATTCCTTCCGCTACATCTCGGACT 202
|||||
467 ThrCysAlaAlaThrThrCysThrCys---CysThrThrThrAlaThrThrCysAla 485
|||||

Qy 203 GGGACCTAGACCTGGCGGTGCTACCGCGTCACCTGGTTCACCTCGTGGAGCCCGCTGCT 262
|||||
486 GlyCys---ThrThrThrCysAlaThrGlyAlaCysGlyCysThrGlyGlyThrCysAla 504
|||||

Qy 263 ACGACTGTGCCGACATGTGCCCGACATTTCGC-----GAGGGGAACCCCAACC 310
|||||
505 ThrThrAla-----ThrCysThrThrGlyCysThrGlyCysAlaThrGlyThrAlaThr 522
|||||

Qy 311 TCAGCTGTGAGGATCTCACCGCGCGCTCTACTTCTGTGAGGACGCAAGGCTGAGCCGG 370
::: ||| |||
523 ThrCysThrGlyGlyGlyCysAlaThrThrGlyThrAlaThrThrThrThrThrThr 542
::: ||| |||

Qy 371 AGGGCGTGGCGCGCTGCACCGCGCCGGGG 400
|||||
543 GlyAlaThrGlyGlyCysThrGlyThrGly 552
|||||

RESULT 9
US-10-184-644-103
; Sequence 103, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:

```

```

: APPLICANT: Baker, Kevin P.
: APPLICANT: Chen, Jian
: APPLICANT: Desnoyers, Luc
: APPLICANT: Goddard, Audrey
: APPLICANT: Godowski, Paul J.
: APPLICANT: Gurney, Austin L.
: APPLICANT: Pan, James
: APPLICANT: Smith, Victoria
: APPLICANT: Watanabe, Colin K.
: APPLICANT: Wood, William I.
: APPLICANT: Zhang, Zemin
: TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
: TITLE OF INVENTION: ACIDS ENCODING THE SAME
: FILE REFERENCE: P3430R1C227
: CURRENT APPLICATION NUMBER: US/10/184,644
: CURRENT FILING DATE: 2002-06-28
: Prior Application removed - See File Wrapper or Palm
: NUMBER OF SEQ ID NOS: 612
: SEQ ID NO 103
: LENGTH: 2558
: TYPE: DNA
: ORGANISM: Homo Sapien
: US-10-184-644-103

Alignment Scores:
Pred. No.: 0.333 Length: 2558
Score: 98.50 Matches: 33
Percent Similarity: 32.81% Conservative: 9
Best Local Similarity: 25.78% Mismatches: 57
Query Match: 8.95% Indels: 29
DB: Gaps: 3

US-09-966-880A-7_COPY_80_676 (1-597) x US-10-184-644-103 (1-2558)
Qy 20 ACCGGAGGAGAGTCTTTTACCAATTCAAAAATGCGCTAAGGCTCGCGGTGAGA 79
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1734 ThrGlyGlyAlaAlaAlaThrGlyAlaThrThrThrThrGlyAlaGly--GlyThrGly 175
||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 80 CCTACCTGTGCTAGCTAGTGAAGAGCGGTGACAGTGTACATCCCTTTTCACGTGGACTTTG 139
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1753 ThrThrCysThrThr-----
||||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

QY 140 GTTATCTTCGAATAAGAACGGCTGCCACGTGGAATTGCTCTCTCCGCTACATCTCGG 199
Db 1758 -----CysCysAlaAlaCysGlyAlaCysThrThrGlyAlaAlaThr 1772
QY 200 ACTGGACCTAGACCTTGGCGGTGCTACCGCGTACCTGGTTCACCTCTGGAGCCCT 259
Db 1773 ThrGlyCysThrThrCysAlaGlyGlyCysAlaGlyCysAlaCysGlyGlyThr 1792
QY 260 GCTACGACTGTGCCCGACATGTGCCGACTTCTGGAGGGAACCCCAACCTCAGTCTGA 319
Db 1793 AlaThrAla-----CysThrAlaAlaAlaAlaAlaAlaAlaAlaThrThr 1804
QY 320 GGATCTTCACCGCGCGCTCTACTCTGTGAGACCGCAAGGCTGAGCCCGAGGCGCTGC 379
Db 1805 GlyGlyGlyAlaAlaAlaCysAlaAlaAlaCysAlaAlaAlaAlaAlaAlaThrThr 1824
QY 380 GCGCGGTGCACCGCGCGGGGTGC 403
Db 1825 GlyGlyCysThrAlaThrCysCys 1832

RESULT 10
US-10-184-634-103
; Sequence 103, Application US/10184634
; Publication No. US20030068684A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P34301C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; PRIOR FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 103
; LENGTH: 2558
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-103

Alignment Scores:
Pred. No.: 0.333 Length: 2558
Score: 98.50 Matches: 33
Percent Similarity: 32.81% Conservative: 9
Best Local Similarity: 25.78% Mismatches: 57
Query Match: 8.95% Indels: 29
DB: 9 Gaps: 3

US-09-966-880A-7_COPY_80_676 (1-597) x US-10-184-634-103 (1-2558)
QY 20 ACCGAGGAAGTTTCTTTACCAATTCAAATAATGTCGCTGGGTAAAGGTGCGCGTGAGA 79
Db 1734 ThrGlyGlyAlaAlaAlaThrGlyAlaThrThrThrGlyAlaGly---GlyThrGly 1752
QY 80 CTTACTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACATCTTTTCACTGGACTTG 139
Db 1753 ThrThrCysThrThr-----1757
QY 140 GTTATCTTCGAATAGAGCGTGCACGTGGAATTGCTCTCTCCGCTACATCTCGG 199
Db 1758 -----CysCysAlaAlaCysGlyAlaCysThrThrGlyAlaAlaThr 1772
QY 200 ACTGGACCTAGACCTTGGCGGTGCTACCGCGTACCTGGTTCACCTCTGGAGCCCT 259
Db 1773 ThrGlyCysThrThrCysAlaGlyGlyCysAlaGlyCysAlaCysGlyGlyThr 1792
QY 260 GCTACGACTGTGCCCGACATGTGCCGACTTCTGGAGGGAACCCCAACCTCAGTCTGA 319
Db 1793 AlaThrAla-----CysThrAlaAlaAlaAlaAlaAlaAlaAlaThrThr 1804
QY 320 GGATCTTCACCGCGCGCTCTACTCTGTGAGACCGCAAGGCTGAGCCCGAGGCGCTGC 379
Db 1805 GlyGlyGlyAlaAlaAlaCysAlaAlaAlaCysAlaAlaAlaAlaAlaAlaThrThr 1824
QY 380 GCGCGGTGCACCGCGCGGGGTGC 403
Db 1825 GlyGlyCysThrAlaThrCysCys 1832

RESULT 11
US-10-184-644-65
; Sequence 65, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P34301C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; PRIOR FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 65
; LENGTH: 1572
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-65

Alignment Scores:
Pred. No.: 0.364 Length: 1572
Score: 97.50 Matches: 35
Percent Similarity: 33.33% Conservative: 10
Best Local Similarity: 25.93% Mismatches: 53
Query Match: 8.86% Indels: 37
DB: 9 Gaps: 4

US-09-966-880A-7_COPY_80_676 (1-597) x US-10-184-644-65 (1-1572)
QY 23 GGAGGAAGTTTCTTTACCAATTCAAATAATGTCGCTGGGTAAAGGTGCGCGTGAGACCT 82
Db 357 GlyGlyAlaGlyGlyThrGlyAlaThrGlyCysAlaGlyThrAlaThrGlyAlaCysThr 376
QY 83 ACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACATCTTTTCACTGGACTTTGGTT 142
Db 377 AlaCys-----378
QY 143 ATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTCTCTCCGCTACATCTCGGACT 202
Db 379 -----ThrAlaThrThrCysThrThrCysAlaThrAlaThrThrThrThr 393
QY 203 GGGACCTAGACCTTGGCGGTGCTACCGCGTACCTGGTTCACCTCTGAGCCCTGCT 262
Db 394 GlyAla---ThrAlaThrAlaThrThrThrCysThrThrCysThrGlyGlyCysAlaGly 412
QY 263 ACGACTGTGCCCGACATGTGCCGACTTCTGGAGGGAACCCCAACCTCAGTCTGAGGA 322
Db 413 ThrThrThr-----ThrThrCysGlyAlaThrThrThrAlaAlaAlaGly 427

```



Db 448 GlyThrGlyThrGlyCysAlaGlyAlaCysThrGlyCysGlyCys 462

RESULT 13

US-10-184-644-319

; Sequence 319, Application US/10184644

; Publication No. US2003004930A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

; APPLICANT: Desnoyers, Luc

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Pan, James

; APPLICANT: Smith, Victoria

; APPLICANT: Watanabe, Colin K.

; APPLICANT: Wood, William I.

; APPLICANT: Zhang, Zemin

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC

; FILE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3430R1C227

; CURRENT APPLICATION NUMBER: US/10/184,644

; CURRENT FILING DATE: 2002-06-28

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 319

; LENGTH: 2103

; TYPE: DNA

; ORGANISM: Homo Saplen

US-10-184-644-319

Alignment Scores:

Pred. No.:	Score:	Length:
0.447	2103	
97.00	35	
Percent Similarity: 36.13%	Conservative: 8	
Best Local Similarity: 29.41%	Mismatches: 64	
Query Match: 8.81%	Indels: 12	
DB:	Gaps: 4	

US-09-966-880A-7\_COPY\_80\_676 (1-597) x US-10-184-644-319 (1-2103)

Qy	56	GCTGGGCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTG	115
Db	1568	AlaGlyCysCysThrGlyThrAlaThrThrCysAlaThr--ThrThrGlyThrThrCys	1586
Qy	116	CTACATCCTTTTCACGTGACATTTGGTTATCTTCGCAATAGAACAGCGTCCACGTGGAAT	175
Db	1587	-----ThrCysThrAlaGlyAlaAlaGlyThrThrThrThrGlyThr	1600
Qy	176	TGCTCTTCCTCCGTACATCTCGACTGGG-----ACCTAGACCTGGCGGTGCT	226
Db	1601	CysAlaGlyAlaAlaThrThrThrThrGlyAlaCysThrThrGlyThrThrGlyAlaCys	1620
Qy	227	ACGCGCTACCTGGTTCCACCTCCTCTGGAGCCCTCTGCTACGACTGTGCCCGACATGTGGCCG	286
Db	1621	AlaThrAlaAlaAlaThrThrThrGlyThrAlaAlaThrGlyCysAlaThrAlaThrAla	1640
Qy	287	ACTTTCGCGAGGGAACCCCAACTCAGTCTGAGGATCTTCACGGCGGCTCTACTTCT	346
Db	1641	ThrAlaCysAlaAlaThrThrThrGlyAlaAlaGlyCysAlaCysThrCysCysThrThr	1660
Qy	347	GTGAGGACCGAAGGCTGAGCCCGGGCGTGGCGGCTGCACCGCGCGGGTG	403
Db	1661	ThrThrCysThrThrCysAlaGlyThrThrCys-----CysThrCysAlaGlyCys	1677

RESULT 14

US-10-184-634-319

; Sequence 319, Application US/10184634

; Publication No. US2003005868A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Chen, Jian

```
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 319
; LENGTH: 2103
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-634-319

Alignment Scores:
Pred. No.: 0.447          Length: 2103
Score: 97.00             Matches: 35
Percent Similarity: 36.13% Conservative: 8
Best Local Similarity: 29.41% Mismatches: 64
Query Match: 8.81%       Indels: 12
DB: 9                    Gaps: 4

US-09-966-880a-7_COPY_80_676 (1-597) x US-10-184-634-319 (1-2103)
QY 56 GCTGGCTAAGGTCGGCTGAGACCTACCTGTGTCTGTAGTGAAGAGCGGTGACAGTG 115
    ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1568 AlaGlyCysCysThrGlyThrAlaThrThrCysAlaThr---ThrThrGlyThrThrCys 1586
QY 116 TCATATCCTTTTCACTGGACTTGTGTTATCTTCGCAATAAGAACGGCTGCCACGTGGGAA 175
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1587 -----ThrCysThrAlaGlyAlaAlaGlyThrThrThrGlyThr 1600
QY 176 TGCTCTTCTCGGTACATCTCGGACTGGG-----ACCTAGACCTGGCGGCTGCT 226
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1601 CysAlaGlyAlaAlaThrThrThrThrThrGlyAlaCysThrThrGlyThrGlyAlaCys 1620
QY 227 ACCGGGTACCTGGTTCACCTCCCTGGAGCCCTCTGTAGACTGTGCCGACATGTGGCG 286
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1621 AlaThrAlaAlaAlaThrThrThrThrThrAlaAlaThrGlyCysAlaThrAlaThrAla 1640
QY 287 ACTTTCGCGAGGAAACCCAACTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCT 346
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1641 ThrAlaCysAlaAlaThrThrThrThrThrGlyAlaAlaGlyCysAlaCysThrCysThr 1660
QY 347 GTGAGGACCGCAGGCTGAGCCCGAGGGGCTGGCGGCTGCACCGCGCGCGGGTGC 403
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1661 ThrThrCysThrThrCysAlaGlyThrThrCys-----CysThrCysAlaGlyCys 1677

RESULT 15
US-10-184-644-39
; Sequence 39, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
```

```
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 39
; LENGTH: 1819
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-39

Alignment Scores:
Pred. No.: 0.609          Length: 1819
Score: 95.50             Matches: 41
Percent Similarity: 38.97% Conservative: 12
Best Local Similarity: 30.15% Mismatches: 62
Query Match: 8.67%       Indels: 21
DB: 9                    Gaps: 6

US-09-966-880a-7_COPY_80_676 (1-597) x US-10-184-644-39 (1-1819)
QY 20 ACCGGAGGAAGTTTCTTTTACCAATTCAAAATGTCGCTGGCTAAGGTCGGGTGAGA 79
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1477 ThrGlyGlyAlaAlaCysGlyAlaGlyThrThrThrGly-----AlaGlyThrAla 1494
QY 80 CCTACCTGTGTCTAGTGTAGAGAGCGGTGACAGTGCTACATCCTTTTCACTGGACTTTG 139
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1495 AlaThrCysAlaGlyGlyAlaAlaGlyThrAla-----ThrAla 1507
QY 140 GTTATCTTCGAATAAGAACGGCTGCCACGTGGGAATGCTCTTCTCGCTACATCTCGG 199
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1508 ThrCysThrAlaThrAlaThrGlyAlaThr-----CysThrThrGlyAlaThrAlaThr 1525
QY 200 ACTGGGACCTAGACCCCTGGGC---GCTGCTACCGCTCACCTGCTTCCACCTCCTGGAGCC 256
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1526 ThrGlyThrThrThrThrAlaThrAlaThrAlaAlaThrThrThrGlyAlaAlaGly 1545
QY 257 CCTGCTACGACTGTGCCCGACATGTGCCCGACTTTCGCGAGGAAACCCCAAGCTCAGTC 316
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1546 ThrCysThrAlaAlaAlaAlaGlyAlaCysThrGlyCysAlaThrThrThrThr--- 1564
QY 317 TGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAGGCTGACCCGAGGGGC 376
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1565 ---AlaAlaAlaCysAlaAlaGlyThrThrAlaGlyThrAlaThrThrAlaThrGly 1583
QY 377 TCGCGC-----GCTGCACCGCGCGGGTGC 403
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1584 CysGlyThrThrGlyCysCysAlaCysGlyThrAlaGlyCys 1599

Search completed: June 14, 2003, 19:09:17
Job time : 38.2225 secs
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2003, 17:59:23 ; Search time 71 Seconds  
(without alignments)  
371.600 Million cell updates/sec

Title: US-09-966-880A-8

Perfect score: 1086

Sequence: 1 MDSLMLNRKFLYQFNVRW.....ILLPLYEVDLDRFRTGL 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1086	100.0	198	21	Human activation-i
2	1008	92.8	198	21	Mouse activation-i
3	390	35.9	384	20	Amino acid sequenc
4	390	35.9	384	22	Human secreted pro
5	390	35.9	384	23	Human polypeptide
6	388	35.7	384	21	Amino acid sequenc
7	369.5	34.0	190	19	Human RNA editing
8	369.5	34.0	190	21	Human RNA editing
9	369.5	34.0	190	22	Human polypeptide
10	369.5	34.0	210	22	Novel human enzyme

11	369.5	34.0	221	22	Human polypeptide
12	369.5	34.0	222	21	Human prostate can
13	369.5	34.0	222	22	Novel human enzyme
14	336.5	31.0	268	23	Human RNA metaboli
15	321.5	29.6	272	22	Novel human enzyme
16	309.5	28.5	163	23	Human secretory po
17	274	25.2	224	23	Human AAD45360 pro
18	274	25.2	224	23	Human APOBEC2 prot
19	271	25.0	222	19	A novel human mRNA
20	262.5	24.2	367	22	Human polypeptide
21	208.5	19.2	261	22	Novel human secret
22	204.5	18.8	169	22	Novel human secret
23	203	18.7	113	21	Human secreted pro
24	199	18.3	236	15	Apo-B RNA editing
25	198	18.2	127	22	Human protein sequ
26	193	17.8	229	15	Apo-B RNA editing
27	103.5	9.5	94	21	Human secreted pro
28	102	9.4	720	22	Drosophila melanog
29	84	7.7	51	22	Peptide #5883 enco
30	84	7.7	51	22	Protein #5554 enco
31	84	7.7	51	22	Human brain expres
32	84	7.7	51	22	Human bone marrow
33	84	7.7	51	22	Peptide #5602 enco
34	84	7.7	51	22	Peptide #5853 enco
35	84	7.7	51	23	Human peptide enco
36	81.5	7.5	2342	22	Drosophila melanog
37	80.5	7.4	410	22	Human transferrase
38	78.5	7.2	402	22	Pseudorabies virus
39	77.5	7.1	261	23	Human caspase-12 i
40	77.5	7.1	402	15	Glycoprotein 50 (g
41	77.5	7.1	402	21	Novel recombinant
42	77.5	7.1	402	22	Pseudorabies virus
43	77.5	7.1	402	22	Pseudorabies virus
44	77.5	7.1	402	22	Pseudorabies virus
45	77.5	7.1	663	22	Human PRO-C-MG.72

ALIGNMENTS

RESULT 1

AAAB24198

ID AAB24198 standard; Protein: 198 AA.

XX AAB24198;

XX 05-FEB-2001 (first entry)

DE Human activation-induced cytidine deaminase SEQ ID NO:8.

XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; antiallergic;  
KW antianemic; antisthmatic; ophthalmological; anti-Hiv; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; IgG subclass selection disorder.

OS Homo sapiens.

XX WO2000058480-A1.

XX 05-OCT-2000.

XX 28-MAR-2000; 2000WO-JP01918.

XX 29-MAR-1999; 99JP-0087192.

XX 24-JUN-1999; 99JP-0178999.

XX 27-DEC-1999; 99JP-0371382.

PA (NIBS) JAPAN TOBACCO INC.  
 XX (HONJ/) HONJO T.  
 PI Honjo T, Muramatsu M;  
 XX WPI; 2000-611715/58.  
 DR N-PSDB; AAC55312.  
 XX  
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 XX  
 XX Claim 1; Page 140-141; 174pp; Japanese.  
 XX  
 CC The present sequence is human activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and  
 CC dermatological activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, allergic  
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
 CC immunodeficiency disorder, MHC (major histocompatibility class)  
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
 CC encoding AID may be used for gene therapy and the antibodies to the AID  
 CC protein may be used for diagnosis and treatment of these disorders.  
 XX  
 XX Sequence 198 AA;  
 SQ  
 Query Match 100.0%; Score 1086; DB 21; Length 198;  
 Best Local Similarity 100.0%; Pred. No. 7.1e-116;  
 Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60  
 DB 1 MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60  
 QY 61 FLRYISDWLDPGRCYRVWTSWSPCYDCARHVAFLRGNPNLSLRIFTARLYFCEDRK 120  
 DB 61 FLRYISDWLDPGRCYRVWTSWSPCYDCARHVAFLRGNPNLSLRIFTARLYFCEDRK 120  
 QY 121 AEPEGLRLRHAGVOIAIMTFKDYFCVWNTFVENHRTFKAWEGHLSVRLSRLRL 180  
 DB 121 AEPEGLRLRHAGVOIAIMTFKDYFCVWNTFVENHRTFKAWEGHLSVRLSRLRL 180  
 QY 181 LPLYEVDLDRDAFRTLGL 198  
 DB 181 LPLYEVDLDRDAFRTLGL 198  
 RESULT 2  
 AAB24197  
 ID AAB24197 standard; Protein; 198 AA.  
 XX  
 XX AAB24197;  
 AC  
 XX 05-FEB-2001 (first entry)  
 DT  
 XX Mouse activation-induced cytidine deaminase SEQ ID NO:2.  
 DE  
 XX  
 XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; antiallergic;  
 KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatologic;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease;

KW auto immunodeficiency syndrome; IgG subclass selection disorder.  
 XX Mus musculus.  
 XX WO200058480-A1.  
 XX 05-OCT-2000.  
 XX 28-MAR-2000; 2000WO-JP01918.  
 XX 29-MAR-1999; 99JP-0087192.  
 PR 24-JUN-1999; 99JP-0178999.  
 PR 27-DEC-1999; 99JP-0371382.  
 XX (NIBS) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 PI Honjo T, Muramatsu M;  
 XX WPI; 2000-611715/58.  
 DR N-PSDB; AAC55307.  
 XX Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 XX  
 XX Claim 1; Page 131-132; 174pp; Japanese.  
 XX  
 CC The present sequence is mouse activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and  
 CC dermatological activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, allergic  
 CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
 CC immunodeficiency disorder, MHC (major histocompatibility class)  
 CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
 CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
 CC encoding AID may be used for gene therapy and the antibodies to the AID  
 CC protein may be used for diagnosis and treatment of these disorders.  
 XX  
 XX Sequence 198 AA;  
 SQ  
 Query Match 92.8%; Score 1008; DB 21; Length 198;  
 Best Local Similarity 92.9%; Pred. No. 6e-107;  
 Matches 183; Conservative 6; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60  
 DB 1 MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60  
 QY 61 FLRYISDWLDPGRCYRVWTSWSPCYDCARHVAFLRGNPNLSLRIFTARLYFCEDRK 120  
 DB 61 FLRYISDWLDPGRCYRVWTSWSPCYDCARHVAFLRGNPNLSLRIFTARLYFCEDRK 120  
 QY 121 AEPEGLRLRHAGVOIAIMTFKDYFCVWNTFVENHRTFKAWEGHLSVRLSRLRL 180  
 DB 121 AEPEGLRLRHAGVOIAIMTFKDYFCVWNTFVENHRTFKAWEGHLSVRLSRLRL 180  
 QY 181 LPLYEVDLDRDAFRTLGL 197  
 DB 181 LPLYEVDLDRDAFRTLGL 197  
 RESULT 3  
 AAY42383  
 ID AAY42383 standard; Protein; 384 AA.  
 XX  
 XX AAY42383;

Qy	116	CEDRKAPGEGLRLRHRCAGVQIAITMTFKDYFCWNTFVENHRTFFKAWBGLHENSVRLSQ	17
Db	316	:-DDGRCQCEGLRTLAEGAKISIMTYSFKHCWDTFVDHQGCPFPWDGLDEHSODLSGR	374
Qy	176	LRRL 180	
Db	375	LRAIL 379	
RESULT 4			
AAU39075	ID	AAU39075 standard; Protein: 384 AA.	
XX	AC	AAU39075;	
XX	DT	2002 (first entry)	
XX	DE	Human secreted protein lp547_4.	
XX	KW	Human; secreted protein; antiinflammatory; immunosuppressive;	
XX	KW	nootropic; neuroprotective; antiarthritic; antimicrobial; vulnerary;	
XX	KW	cytostatic; antidiabetic; virucide; antinfertility; anticonvulsant;	
XX	KW	vasotropic; antiparkinsonian; immunostimulant; dermatological;	
XX	KW	antirheumatic; anticumor; antiulcer; osteopathic; tranquiliser;	
XX	KW	cerebroprotective; cytokine; cell proliferation; cell differentiation;	
XX	KW	immune deficiency; severe combined immunodeficiency; SCID; tumour;	
XX	KW	autoimmune disorder; multiple sclerosis; rheumatoid arthritis;	
XX	KW	graft-versus-host disease; myeloid deficiency; wound healing; ulcer;	
XX	KW	periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;	
XX	KW	Parkinson's disease; Huntington's disease; infection; cardiac disease;	
XX	KW	stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;	
XX	KW	food supplement; vaccine.	
XX	OS	Homo sapiens.	
XX	PN	WO200175068-A2.	
XX	PD	11-OCT-2001.	
XX	PF	22-MAR-2001; 2001WO-US09369.	
XX	PR	30-MAR-2000; 2000US-0539330.	
XX	PR	04-DEC-2000; 2000US-0729674.	
XX	PA	(GEMV ) GENETICS INST INC.	
XX	PI	Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;	
XX	PI	Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;	
XX	PI	Clark H, Fechtel K, Merberg D;	
XX	DR	WPI; 2001-639363/73.	
XX	DR	N-PSDB; AAS59293.	
XX	PT	Secreted human proteins, useful as vaccine for treating various	
XX	PT	diseases such as autoimmune disorders (e.g. multiple sclerosis), and	
XX	PT	nervous system disorders (e.g. stroke) -	
XX	PS	Disclosure; Page 580-581; 619pp; English.	
XX	CC	The invention relates to novel human secreted proteins, the nucleic	
XX	CC	acids encoding them. The protein may exhibit cytokine, cell proliferation	
XX	CC	or cell differentiation activity or may induce production of other	
XX	CC	cytokines in certain cell populations and may exhibit immune stimulating	
XX	CC	or immune suppressing activity, which is useful for the treatment of	
XX	CC	various immune deficiencies and disorders e.g. severe combined	
XX	CC	immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,	
XX	CC	systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary	
XX	CC	inflammation. The proteins are also useful in the treatment of diseases	
XX	CC	and disorders including tissue, skin and organ transplantation and in	
XX	CC	graft-versus-host diseases (GVHD), in the induction of tumour immunity,	
XX	CC	myeloid or lymphoid cell deficiencies, wound healing and tissue repair,	
XX	CC	in the treatment of burns, incisions and ulcers; as well as in treatment	
XX	CC	of periodontal disease, osteoporosis or osteoarthritis, mediated by	

US2001039335-A1.

08-NOV-2001.

04-DEC-2000; 2000US-0729674.

10-APR-1997; 97US-126425P.

04-DEC-1997; 97US-067454P.

20-DEC-1997; 97US-068379P.

02-JAN-1998; 98US-070346P.

07-JAN-1998; 98US-070643P.

08-JAN-1998; 98US-070755P.

13-JAN-1998; 98US-071304P.

22-JAN-1998; 98US-072134P.

30-JAN-1998; 98US-073095P.

18-FEB-1998; 98US-075038P.

30-MAR-2000; 2000US-0539330.

The invention relates to isolated polynucleotides (ABA90876-ABA90968 and ABA90980) and encoded proteins (ABA55698-AB557800), especially polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and proteins SEQ ID NO 2 (AB55698) and SEQ ID NO 20 (AB55707), contained in clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1 are deposited with the American Type Culture Collection (ATCC) under accession number 98599. The polynucleotides and encoded polypeptides have a cytostatic, anti-inflammatory, immunomodulator, vulnerary, neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic and anti-inflammatory activity and acting as cytokine modulators, hematopoiesis regulators, tissue growth modulators and/or cadherin suppressors. The polypeptides and polynucleotides are useful in gene therapies, particularly for preventing, treating or ameliorating any of the following diseases: immune deficiency and disorders; e.g. bacterial or fungal infections, autoimmune disorders, cancer, systemic lupus erythematosus or graft-versus-host disease; myeloid or lymphoid cell deficiencies; wound, burns, incisions and ulcers, osteoporosis or osteoarthritis; central and peripheral nervous system diseases and neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis or Shy-Drager syndrome; haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis or systemic inflammatory response syndrome, ischaemia-reperfusion injury, endotoxin lethality, arthritis, inflammatory bowel disease or Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus foliaceus.

Seq	Sequence	384 AA:	Query Match	35.9%:	Score 390:	DB 23:	Length 384:
			Best Local Similarity	44.9%:	Pred. No. 7 5e-36:		
			Matches 83:	Conservative 31:	Mismatches 59:	Indels 12:	Gaps 3:
QY	6	MNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLD--FGVLRNK-----NGC 55					
Db	197	MDPTFTFNPNPWRGRHETLYCYEVERHMDTVLLNQRRGFLCNOAPHKHGFLSGR 256					
QY	56	HVELLFTRYISDWLDGPGRCYRVYTFWTSKSPCYDCARHVADFLRGNPNLSLRIFTARLYF 115					
Db	257	HAELCFLDVLPFNKLLDQDYRYTCFTSWSPCESCAQEMAKFISNKKHVSLCIFTARY- 315					
QY	116	CDRKAEPGLRLRLRACGVIAITMKDYFYCWNTFVENHRTFKAMECLLHNSVRLSRQ 175					
Db	316	-DQGRQCEGLRLTAEAGAKISIMTYSEFKHCWDTFVDHQGCFPGPDWDLDEHSODLSGR 374					







```
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX WPI; 2001-442253/47.
DR N-PSDB; AAI58061.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
PT
XX Example 3; SEQ ID NO 2050; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 190 AA;
SQ
Query Match 34.0%; Score 369.5; DB 22; Length 190;
Best Local Similarity 44.9%; Pred. No. 6.3e-34;
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;
QY 11 FLYQKNRWKGRRETYLCYVVK--RRDSATSFSLDFGLRN---KNGCHVELLFLRY 64
DB 17 FFFQKNLWEANDRNETWLCFTVEGIKRKSVMWSKT--GVFRNQVDSETHCHAEKFLSW 74
QY 65 ISDWLDLPGRCYRVTFWFTSWSPCYDCARHVAFLRGPNLNLRIFTARLYFCEDRKAEP 124
DB 75 FCDLILSPNTKYQVTWTSWSPCDGAGEVAEFLARHSNVNLFTIFARLYYFQ-YPCYQE 133
QY 125 GLRRLHRAQVQIAITFKDYFCWNTFVENHRTFKAWGLHENSIVLSRLRL 180
DB 134 GLRSLSQEGVAVEIMDYEDFKYCWENFVNDNEPFPKWLKLTNFKLLKRLRESL 189
RESULT 10
AAU23550
ID AAU23550 standard; Protein; 210 AA.
XX
XX AAU23550;
XX
XX 17-DEC-2001 (first entry)
XX
XX Novel human enzyme polypeptide #636.
XX
XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
XX Homo sapiens.
XX
XX WO200155301-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
PR
PR 04-FEB-2000; 2000US-0180628.
PR
PR 24-FEB-2000; 2000US-0184664.
PR
PR 02-MAR-2000; 2000US-0186350.
PR
PR 16-MAR-2000; 2000US-0189874.
PR
PR 17-MAR-2000; 2000US-0190076.
PR
PR 18-APR-2000; 2000US-0198123.
PR
PR 19-MAY-2000; 2000US-0205515.
PR
PR 07-JUN-2000; 2000US-0209467.
PR
PR 28-JUN-2000; 2000US-0214886.
PR
PR 30-JUN-2000; 2000US-0215135.
PR
PR 07-JUL-2000; 2000US-0216647.
PR
PR 07-JUL-2000; 2000US-0216880.
PR
PR 11-JUL-2000; 2000US-0217487.
PR
PR 11-JUL-2000; 2000US-0217496.
PR
PR 14-JUL-2000; 2000US-0218290.
PR
PR 26-JUL-2000; 2000US-0220963.
PR
PR 26-JUL-2000; 2000US-0220964.
PR
PR 14-AUG-2000; 2000US-0224518.
PR
PR 14-AUG-2000; 2000US-0224519.
PR
PR 14-AUG-2000; 2000US-0225213.
PR
PR 14-AUG-2000; 2000US-0225214.
PR
PR 14-AUG-2000; 2000US-0225266.
PR
PR 14-AUG-2000; 2000US-0225267.
PR
PR 14-AUG-2000; 2000US-0225268.
PR
PR 14-AUG-2000; 2000US-0225270.
PR
PR 14-AUG-2000; 2000US-0225447.
PR
PR 14-AUG-2000; 2000US-0225757.
PR
PR 14-AUG-2000; 2000US-0225758.
PR
PR 14-AUG-2000; 2000US-0225759.
PR
PR 18-AUG-2000; 2000US-0226279.
PR
PR 22-AUG-2000; 2000US-0226681.
PR
PR 22-AUG-2000; 2000US-0226868.
PR
PR 23-AUG-2000; 2000US-0227009.
PR
PR 30-AUG-2000; 2000US-0228924.
PR
PR 01-SEP-2000; 2000US-0229287.
PR
PR 01-SEP-2000; 2000US-0229344.
PR
PR 01-SEP-2000; 2000US-0229345.
PR
PR 05-SEP-2000; 2000US-0229509.
PR
PR 05-SEP-2000; 2000US-0229513.
PR
PR 06-SEP-2000; 2000US-0230437.
PR
PR 06-SEP-2000; 2000US-0230438.
PR
PR 08-SEP-2000; 2000US-0231242.
PR
PR 08-SEP-2000; 2000US-0231243.
PR
PR 08-SEP-2000; 2000US-0231244.
PR
PR 08-SEP-2000; 2000US-0231413.
PR
PR 08-SEP-2000; 2000US-0231414.
PR
PR 08-SEP-2000; 2000US-0232080.
PR
PR 08-SEP-2000; 2000US-0232081.
PR
PR 12-SEP-2000; 2000US-0231968.
PR
PR 14-SEP-2000; 2000US-0232397.
PR
PR 14-SEP-2000; 2000US-0232398.
PR
PR 14-SEP-2000; 2000US-0232399.
PR
PR 14-SEP-2000; 2000US-0232400.
PR
PR 14-SEP-2000; 2000US-0232401.
PR
PR 14-SEP-2000; 2000US-0233063.
PR
PR 14-SEP-2000; 2000US-0233064.
PR
PR 21-SEP-2000; 2000US-0233065.
PR
PR 21-SEP-2000; 2000US-0234223.
PR
PR 21-SEP-2000; 2000US-0234274.
PR
PR 25-SEP-2000; 2000US-0234997.
PR
PR 25-SEP-2000; 2000US-0234998.
PR
PR 26-SEP-2000; 2000US-0235484.
PR
PR 27-SEP-2000; 2000US-0235834.
PR
PR 27-SEP-2000; 2000US-0235836.
PR
PR 29-SEP-2000; 2000US-0236327.
PR
PR 29-SEP-2000; 2000US-0236367.
PR
PR 29-SEP-2000; 2000US-0236368.
```



PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX (HYSE-) HYSEQ INC.  
 PA  
 XX  
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI; 2001-442253/47.  
 DR N-PSDB; AAI59847.  
 XX  
 XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 XX Example 2; SEQ ID NO 5622; 10078pp; English.  
 XX  
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with neurotropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukaemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 XX Sequence 221 AA;

Query Match 34.0%; Score 369.5; DB 22; Length 221;  
 Best Local Similarity 44.9%; Pred. No. 7.9e-34;  
 Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;  
 QY 11 FLYQFKNVWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN---KNGCHVELLFLRY 64  
 Db 48 FYFOFKNLWEANDRNETWLCFTVEGIKRRSVVSWKT--GVFRNQVDSETHCAERCFLSW 105  
 QY 65 ISDWDLDPGRCYRVWFTSWSPCYDCARHVADEFLRGNPNLSLRIFTARLYFCEDRKAEP 124  
 Db 106 FCDLILSPNTKYQVTWYTSWSPCDGAGEVAEFLARHSNVNLTIFTARLYFQ-YPCYQE 164  
 QY 125 GLRLRHAGVQIAIMTFKDYFCYCNWTFVFNHRTFFKAWEGHLSNVLSQLRRIL 180  
 Db 165 GLRSLSQGVAVEIMDYEDFKYCNWFVNDNPFKPKGLKTNFRLLKRLRESL 220

RESULT 12  
 AAB57061  
 ID AAB57061 standard; Protein: 222 AA.  
 XX  
 XX AAB57061;  
 XX  
 XX 13-MAR-2001 (first entry)  
 XX  
 XX Human prostate cancer antigen protein sequence SEQ ID NO:1639.  
 XX  
 XX Human; prostate cancer; antigen; detection; diagnosis;  
 KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;  
 KW vulnary; gastrointestinal; nephrotropic; antiinfective; gynaecological;  
 KW antibacterial; gene therapy; neural; immune; reproductive; renal;  
 KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;

KW wound; infectious disease.  
 XX  
 XX Homo sapiens.  
 XX WO2000055174-A1.  
 XX 21-SEP-2000.  
 XX  
 XX 08-MAR-2000; 2000WO-US05988.  
 XX  
 XX 12-MAR-1999; 99US-0124270.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.  
 XX  
 XX Rosen CA, Ruben SM;  
 PI  
 XX WPI; 2000-587513/55.  
 DR N-PSDB; AAF16264.  
 XX  
 XX Prostate cancer associated gene sequences, referred to as prostate  
 PT cancer antigens, useful for treatment, prevention, and diagnosis of  
 PT disorders such as prostate cancer -  
 XX  
 XX Claim 11; Page 2097-2098; 2338pp; English.  
 XX  
 XX AAF15566 to AAF16505 encode the human prostate cancer associated  
 CC proteins, called prostate cancer antigens, given in AAB56363 to AAB57302.  
 CC The prostate cancer antigens can have neuroprotective, cytostatic,  
 CC cardioactive, immunomodulatory, muscular, vulnary, gastrointestinal,  
 CC nephrotropic, antiinfective, gynaecological and antibacterial activities,  
 CC and can be used in gene therapy. The prostate cancer antigen  
 CC polynucleotides may be used for detection of prostate cancer, chromosome  
 CC identification, as chromosome markers, and for numerous other diagnostic  
 CC or research purposes. The prostate cancer antigens may be used to treat  
 CC disorders such as neural, immune, muscular, reproductive,  
 CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative  
 CC disorders, wounds, and infectious diseases. AAF16506 to AAF16514 to  
 CC AAB57303 represent sequences used in the exemplification of the present  
 CC invention.

QY 11 FLYQFKNVWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN---KNGCHVELLFLRY 64  
 Db 49 FYFOFKNLWEANDRNETWLCFTVEGIKRRSVVSWKT--GVFRNQVDSETHCAERCFLSW 106  
 QY 65 ISDWDLDPGRCYRVWFTSWSPCYDCARHVADEFLRGNPNLSLRIFTARLYFCEDRKAEP 124  
 Db 107 FCDLILSPNTKYQVTWYTSWSPCDGAGEVAEFLARHSNVNLTIFTARLYFQ-YPCYQE 165  
 QY 125 GLRLRHAGVQIAIMTFKDYFCYCNWTFVFNHRTFFKAWEGHLSNVLSQLRRIL 180  
 Db 166 GLRSLSQGVAVEIMDYEDFKYCNWFVNDNPFKPKGLKTNFRLLKRLRESL 221

RESULT 13  
 AAU23537  
 ID AAU23537 standard; Protein: 222 AA.  
 XX  
 XX AAU23537;  
 XX  
 XX 17-DEC-2001 (first entry)  
 XX  
 XX Novel human enzyme polypeptide #623.  
 XX  
 XX Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
 KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
 KW autoimmune disorder; neurological disorder; metabolic disorder;

KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;  
XX nephrotropic; anticoagulant.  
OS Homo sapiens.  
PN WO200155301-A2.  
XX 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01239.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-02331968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-02334223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.

PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-465566/50.  
DR N-PSDB; AAS41407.  
XX Novel polypeptides and polynucleotides useful for diagnosing,  
PT preventing, treating neural, immune system, muscular, reproductive,  
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
PT diseases  
XX  
XX Claim 11; SEQ ID No 1533; 1180pp; English.  
XX  
XX The present invention relates to the isolation of novel human enzyme  
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences  
CC encoding them. The enzyme polypeptides of the invention may comprise the  
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
CC isomerases or ligases. The sequences of the invention are useful in the  
CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
CC disorders, including hyperproliferative disorders (e.g. cancer),  
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders  
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
CC blood-related disorders (e.g. haemophilia), reproductive disorders  
CC (e.g. infertility) and infectious disorders (e.g. influenza). The  
CC polynucleotides of the invention can also be used in gene therapy.  
CC AAU22915-AAU223814 represent the novel human enzyme polypeptides of the  
CC invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 222 AA;  
SQ  
Query Match 34.0%; Score 369.5; DB 22; Length 222;  
Best Local Similarity 44.9%; Pred. No. 7.9e-34;  
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;  
QY 11 FLYQFKNVWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----KNGCHVELLFLRY 64  
Db 49 FYFQKNLWEANDRNTELCTFVEGIKRSVWSMKT--GVFRNQVDSETHCHAEKFLSW 106  
QY 65 ISDWLDPCRCYRVWFTSWSPCYDCARHVAADPLRGNPNLSLRIFTARLYFCEDRAEGL 124  
Db 107 FCDLILSPNTKYQVWWTYSWSPCPDCAGEVAEFLARHSNVNLTFTARLYFQ-YPCYQE 165  
QY 125 GLRRLHRAGVQIAIMTFKDYFCWNTFVENHRTFKAWEGLHENSRLSRLRRL 180  
Db 166 GLRSLSQEGAVAEIMDYEDFKYCWENFVYNDNPFKPKWGLKTNFRLKRLRESL 221  
RESULT 14  
AAE15256  
ID AAE15256 standard; Protein; 268 AA.  
XX  
AC AAE15256;  
XX  
XX 07-MAR-2002 (first entry)  
XX Human RNA metabolism protein-19 (RMEP-19).  
XX Human; RNA metabolism protein-19; RMEP-19; gout; nervous system disorder;  
KW autoimmune; inflammatory; cell proliferative; developmental; thyroiditis;  
KW gene therapy; epilepsy; dementia; stroke; Alzheimer's disease; amnesia;  
KW Parkinson's disease; prion disease; insomnia; endocrine disorder; AIDS;  
KW Acquired Immune Deficiency Syndrome; mental disorder; allergy; anaemia;  
KW asthma; atherosclerosis; Crohn's disease; rheumatoid arthritis; vaccine;  
KW glomerulonephritis; multiple sclerosis; diabetes mellitus; osteoporosis;

KW cancer; cirrhosis; hepatitis; psoriasis; transgenic animal; antitumor;  
KW tranquilliser; drug screening; pancreatitis; renal tubular acidosis;  
KW systemic lupus erythematosus; colitis; ss.  
XX  
XX Homo sapiens.  
XX  
XX Key Location/Qualifiers  
FT Peptide 1..21  
FT /label= Signal\_peptide  
FT Protein 22..268  
FT /label= Mature\_RMEP\_19\_protein  
XX  
XX WO200183524-A2.  
XX  
XX 08-NOV-2001.  
XX  
XX 27-APR-2001; 2001WO-US13862.  
XX  
XX 28-APR-2000; 2000US-200184P.  
PR 04-MAY-2000; 2000US-201875P.  
PR 04-MAY-2000; 2000US-202090P.  
PR 06-JUN-2000; 2000US-210232P.  
PR 25-JUL-2000; 2000US-220553P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX Lal P, Yue H, Tang YT, Lu DAM, Azimzai Y, Au-Young J, Hillman JL;  
PI Baughn MR, Yao MG, Burford N, Batra S, Policky JJ;  
XX  
XX WPI; 2002-034502/04.  
DR N-PSDB; AAD24392.  
XX  
XX New human RNA metabolism protein for diagnosing or treating nervous  
PT system disorders, autoimmune/inflammatory disorders, cell proliferative  
PT disorders and developmental disorders  
XX  
XX Claim 1; Page 145-146; 196pp; English.  
XX  
XX The invention relates to human RNA metabolism proteins (RMEP) and their  
CC corresponding cDNA molecules. RMEP and its DNA are used for diagnosing,  
CC treating and preventing nervous system disorders (epilepsy, dementia,  
CC stroke, Alzheimer's disease, Huntington's disease, Parkinson's disease);  
CC prion diseases; fatal familial insomnia, nutritional and metabolic  
CC diseases of the nervous system; inherited, metabolic, endocrine and  
CC toxic myopathy; mental disorders (mood, anxiety, schizophrenic disorders)  
CC amnesia and Tourette's disorder; autoimmune/inflammatory disorders (AIDS-  
CC acquired immune deficiency syndrome, allergies, anaemia, asthma, gout,  
CC atherosclerosis, Crohn's disease, diabetes mellitus, glomerulonephritis,  
CC Hashimoto's thyroiditis, multiple sclerosis, osteoarthritis, rheumatoid  
CC arthritis, osteoporosis, pancreatitis, systemic lupus erythematosus,  
CC ulcerative colitis, and infections); cell proliferative disorders (cancer  
CC arteriosclerosis, cirrhosis, hepatitis, psoriasis); and developmental  
CC disorders (renal tubular acidosis). RMEP DNA is useful in drug screening  
CC techniques, gene therapy and for creating transgenic animals. The present  
XX sequence is human RMEP-19 protein.  
SQ  
Sequence 268 AA;  
Query Match 31.0%; Score 336.5; DB 23; Length 268;  
Best Local Similarity 40.8%; Pred. No. 6.1e-30;  
Matches 71; Conservative 27; Mismatches 71; Indels 5; Gaps 2;  
QY 11 FLYQFKNVWAKGRRETYLCYVVKRRDSATSFSLDFGYLRN----KNGCHVELLFLRYIS 66  
Db 95 FYFHKNLKACGRNESWLCFTMTVTKHHSVFRKGVFRNQVDSETHCHAEKFLSWFC 154  
QY 67 DWLDLPCRCYRVWFTSWSPCYDCARHVAADPLRGNPNLSLRIFTARLYFCEDRAEGL 126  
Db 155 DDILSPNTNVEYTWYTSWSPCECAGEVAEFLARHSNVNLTFTARLYCFWDTDYQ-BGL 213  
QY 127 RRLHRAGVQIAIMTFKDYFCWNTFVENHRTFKAWEGLHENSRLSRLRRL 180  
Db 214 CSLSQEGASVKIMGYKDFVSCWKNFYVSDDEPFKPKWGLQTNFRLKRLRESL 267

RESULT 15  
AAU23799  
ID AAU23799 standard; Protein; 272 AA.  
XX AC AAU23799;  
XX DT 18-DEC-2001 (first entry)  
XX DE Novel human enzyme polypeptide #885.  
XX KW Human: oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;  
KW ligase; hyperproliferative disorder; immunodeficiency disorder;  
KW autoimmune disorder; neurological disorder; metabolic disorder;  
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;  
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;  
KW nephrotropic; anticoagulant.  
XX OS Homo sapiens.  
XX PN WO200153301-A2.  
XX PD 02-AUG-2001.  
XX PF 17-JAN-2001; 2001WO-US01239.  
XX PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 18-AUG-2000; 2000US-0225759.  
PR 22-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.

PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
XX PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-465566/50.  
DR N-PSDB; AAS41669.  
XX  
PT Novel polypeptides and polynucleotides useful for diagnosing,  
PT preventing, treating neural, immune system, muscular, reproductive,  
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
PT diseases -  
XX  
XX Claim 11; SEQ ID No 1795; 1180pp; English.  
PS  
XX The present invention relates to the isolation of novel human enzyme  
CC polypeptides, and the cDNA (AAS40785-AAS41684) and genomic sequences  
CC encoding them. The enzyme polypeptides of the invention may comprise the  
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
CC isomerases or ligases. The sequences of the invention are useful in the  
CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
CC disorders including hyperproliferative disorders (e.g. cancer),  
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders  
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
CC blood-related disorders (e.g. haemophilia), reproductive disorders  
CC (e.g. infertility) and infectious disorders (e.g. influenza). The  
CC polynucleotides of the invention can also be used in gene therapy.  
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the  
CC invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX SQ Sequence 272 AA;  
Query Match 29.6%; Score 321.5; DB 22; Length 272;  
Best Local Similarity 45.3%; Pred. No. 3.2e-28;  
Matches 68; Conservative 22; Mismatches 51; Indels 9; Gaps 4;  
QY 11 FLYQFNVRWAKGRRETYLCYVVK--RRDSATSFSLDFGLRN----KNGCHVELFLRY 64  
DB 49 FYFQFKNLWEANDRNETWLCFTVEGIKRSVYSWKT--GVFRNQVDSETHCAERCFLSW 106  
QY 65 ISDWDLDPGRCYRVTFWTSWSPCYDCARHVADFLRGPNPLSLRIFTARLYFCEDRKAEP 124  
DB 107 FCDLILSPNTKQVWTWYTSWSPCPDAGEVAEFLARHSNVLNLTIFTARLYFYQ-YPCYOE 165  
QY 125 GLRRLHRAGVQIAIMTFKDYFCWNTFVEN 154  
DB 166 GLRSLSQEGVAEIMDYEDFKYCWFNVYN 195





Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	245	22.6	116	2	G01233		phorbol I - huma
2	225	20.7	236	2	A53853		apolipoprotein B m
3	216.5	19.9	236	2	I59323		apolipoprotein B m
4	211	19.4	229	2	JC4269		apolipoprotein B m
5	211	19.4	229	2	I48249		apolipoprotein B m
6	198	18.2	229	2	I59577		apolipoprotein B m
7	97.5	9.0	369	2	S63464		hypothetical prote
8	81.5	7.5	242	2	T13412		hypothetical prote
9	80	7.4	476	2	C64601		hypothetical prote
10	78.5	7.2	817	2	S53921		fucosyltransferase
11	77.5	7.1	251	2	T32318		hypothetical prote
12	77.5	7.1	328	2	T30999		hypothetical prote
13	77.5	7.1	402	1	VGBE50		glycoprotein D pre
14	77.5	7.1	618	2	G59432		MacGAP protein lim
15	77	7.1	1829	2	T14280		RWL protein - mous
16	76.5	7.0	466	1	S15280		phosphogluconate d
17	75.5	7.0	382	2	A47477		hyaluronidase - ho
18	75.5	7.0	533	2	A47143		retinal pigment m
19	75	6.9	320	2	E70355		conserved hypothet
20	75	6.9	617	2	S52797		mufI protein - hum
21	74.5	6.9	229	2	F97193		single-strand DNA-
22	74	6.8	319	2	A34355		UDP-galactose-4-epi
23	74	6.8	707	2	S78538		site-specific reco
24	73.5	6.8	337	2	A90846		Hnr protein limpo
25	73.5	6.8	337	2	H85703		Hnr protein limpo
26	73.5	6.8	337	2	A36871		37K regulator resp
27	73.5	6.8	483	1	DESHG		phosphogluconate d
28	73.5	6.8	727	2	A04749		1,4-alpha-glucan b
29	72.5	6.7	191	2	S57642		interferon precurs



	Query Match	7.4%	Score 80;	DB 2;	Length 476;
	Best Local Similarity	20.8%;	Pred. No. 4.5;		
	Matches 43;	Conservative 21;	Mismatches 57;	Indels 86;	Gaps 10
Qy	7	NRKFLYQFKVNRWAKGRREYLCYVVKRRDSATSFSLDFGLYLRNKGCHVLLFLRYIS	66		
		:       :	:	:     :	:
Db	235	NKSEFLSYQK-----FNLCF-----ENSCQGYGVTEK-----ILDAYS	258		
Qy	67	D-----W-----DLDPGRCYRVTWTFWSWSPCYDCARHV-----ADFLRGNPNLSLR	107		
Db	259	HTIPIYMGSPSAKDFNPKSFVNVHDNFNDAIDYIKYLTHHPNAYLDMLYENP---LN	315		

```
QY 108 IFARLYFCEDRKAEPEGLRLHRAGVOIAMTFKDYFYCWNTFVEN-----HERTEKAW 163
Db 316 TLGKAYEQD-----LSFKKILDFEKTILENDTIYHNNPFIFYR 355
QY 164 GLHENSIVLSQRLRRILLPIYEVDDL 190
Db 356 DLHE-----PLISIDDL 368

RESULT 10
S53921
hypothetical protein YGR198w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein G7594
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C:Accession: S53921; S59183; S64520; S64516
R:Nombela, C.
submitted to the EMBL Data Library, November 1994
A:Reference number: S53918
A:Accession: S53921
A:Molecule type: DNA
A:Residues: 1-817 <NOM>
A:Cross-references: EMBL:X82775; NID:g791126; PIDN:CAA58017.1; PID:g791130
R:Guereiro, P.; Silva, A.M.E.; Barreiros, T.; Arroyo, J.; Garcia-Gonzalez, M.; Garcia-S
Yeast 11, 1087-1091, 1995
A:Title: The complete sequence of a 9000 bp fragment of the right arm of Saccharomyces c
A:Reference number: S59180; MUID:96076633; PMID:7502584
A:Accession: S59183
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-817 <GUR>
A:Cross-references: EMBL:X82775; NID:g791126; PIDN:CAA58017.1; PID:g791130
R:Guereiro, P.; Barreiros, T.; Cyrne, L.; Soares, H.; Maia e Silva, A.; Rodrigues-Pousa
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64517
A:Accession: S64520
A:Molecule type: DNA
A:Residues: 1-817 <GUW>
A:Cross-references: EMBL:X82775; NID:g1323352; PIDN:CAA97225.1; PID:e243737; PID:g132335
R:Arroyo, J.; Garcia-Gonzalez, M.; Garcia-Saez, M.I.; Sanchez-Perez, M.; Nombela, C.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64499
A:Accession: S64516
A:Molecule type: DNA
A:Residues: 1-240 <ARR>
A:Cross-references: EMBL:X82775; MIPS:YGR198w
C:Genetics:
A:Map position: 7R

Query Match 7.2%; Score 78.5; DB 2; Length 817;
Best Local Similarity 21.3%; Pred. No. 12;
Matches 42; Conservative 33; Mismatches 63; Indels 59; Gaps 9;
QY 41 SFSLDGFLRNKNGCHVELLFLRYISDWLDGRCYRVYTWFTSWSPCYDCARHVADELRG 100
Db 413 SESLD--WLENSTRC-----YVPGNISKVL-TNAWSTLYEIRKYOLDPLVS 455
QY 101 NPNLRLIFRATFYFCEDRK-----APEGLRL-----HRAGVQIAIMTF 141
Db 456 N-NLTSVLCNAMLSSTKEKNADNADVEQEGEEKALRELQFKYSYTLAQORHIEATKTL 514
QY 142 K-----DYFCWNTFV-----ENHERTEKAW-----EGLHENSIVLSQRLRI 179
Db 515 ESILSKNPNYKAWHLLALCRSVQEDKMSYKIVCSVLEAMNESLQNNTLLNDRWQFI 574
QY 180 LLPLYEVDDLRAFRTL 196
Db 575 HLKLTALALIEIEFTGL 591

us-09-966-880a-8.rpr
Page 4
```

Db 180 -----FCKLEA-AKGRRLHRNKRSSQVTRSVQVLF--- 216  
QY 144 YFYCWTFVFNHRTFKAWGLHNSVRLSKQLRRLPL-----YEV 186  
Db 217 HLLCWSPF-----WLENLFAIFR--VRITQLMRIIVNIIHLPPYNCALNPVLYAYRA 269  
QY 187 DLLRDAPRTL 196  
Db 270 ENERTAFKSL 279

## RESULT 13

VGBE50

glycoprotein D precursor - suid herpesvirus 1 (strain Rice)

C:Species: suid herpesvirus 1

A:Note: host Sus scrofa domestica (domestic pig)

C:Date: 31-Mar-1988 #sequence\_revision 31-Mar-1988 #text\_change 16-Jul-1999

C:Accession: A27788

R:Petrovskis, E.A.; Timmins, J.G.; Armentrout, M.A.; Marchiolli, C.C.; Yancey Jr., R.J.;

J. Virol. 59, 216-223, 1986

A:Title: DNA sequence of the gene for pseudorabies virus gp50, a glycoprotein without N-

A:Reference number: A27788; MUID:86281819; PMID:3016293

A:Accession: A27788

A:Molecule type: DNA

A:Residues: 1-402 &lt;PET&gt;

A:Cross-references: GB:M14001; NID:g334051; PIDN:AAC35203.1; PID:g334052

C:Superfamily: herpesvirus glycoprotein D

C:Keywords: glycoprotein; transmembrane protein

F:1-17/Domain: signal sequence #status predicted &lt;SIG&gt;

F:18-402/Product: glycoprotein D #status predicted &lt;GPD&gt;

F:362-378/Domain: transmembrane #status predicted &lt;TMN&gt;

Query Match 7.1%; Score 77.5; DB 1; Length 402;  
Best Local Similarity 24.9%; Pred. No. 6.6;  
Matches 42; Conservative 17; Mismatches 73; Indels 37; Gaps 6;

QY 24 RRTYLCLYVVKRDSATSFSLDFGLYLNKNGCHVELFLRYISDWLDP---GRCYRVT 79  
Db 88 RRPTYRAHV-----AWYRIADGCAHLLYFIEVA---DCDPQVFGRCRRRT 130  
QY 80 WFTSWSPCYDCARHVADFLRGNNPLSLRIETARLYFCEDRKAEPEGLRLRHAGVQIA-- 137  
Db 131 TPMWWTSPADYMFPTDEL-----GLLMVAPGRFNEGQYRLVSDGVNLTLD 178  
QY 138 -IMTFKDYFYCWTFVFNHRTFKAWGLHNSVRLSKQLRRLPLLYE 185  
Db 179 FMVALPEGQCEPVARVDQH-RTYKFGACWSDDSFKRGVDVWRFLTPFIQ 226

## RESULT 14

G59432

MacGAP protein [imported] - huam

C:Species: Homo sapiens (man)

C:Date: 03-Jun-2002 #sequence\_revision 03-Jun-2002 #text\_change 02-Aug-2002

C:Accession: G59432

R:Uchida, T.; Kuramasu, A.; Okumura, K.; Nakao, A.; Ogawa, H.; Ra, C.

submitted to GenBank, December 2001

A:Description: Molecular Cloning and Characterization of A novel GTPase Activating Prote

A:Reference number: G59432

A:Accession: G59432

A:Molecule type: mRNA

A:Residues: 1-618 &lt;UCH&gt;

A:Cross-references: GB:NP\_277050; PID:g15723376; PIDN:NP\_277050.1

Query Match 7.1%; Score 77.5; DB 2; Length 618;  
Best Local Similarity 22.5%; Pred. No. 11;  
Matches 42; Conservative 36; Mismatches 94; Indels 15; Gaps 6;

QY 16 KNRVWAKGRRETYLCYVVKRDSATSFSLDFGLYLNKNGCHVELFLRYISD---WDLDP 72  
Db 204 ETKQSKGDDATLPSFLRPKDKTGTTRIGLAPQDMKKVKCHLALIELTALYDVLGIELQ 263

QY 73 GRCYRVTFWTSWSPCYDCARHVADFLRGNNPLSL-RIETARLYFCEDRKAEPEGLRLHR 131  
Db 264 QKAVKIKTKDSGLFCVPLTALLEQDQKRVPGMRIPLIQFKLISRTEERGLETEGLRIPG 323  
QY 132 AGVQIAIMTFKDYFYCWTFVFNHRTFKAWGL--HNSVRLSKQLRRLPLLYEVDL 189  
Db 324 AAIRIKNL-----CQLEAKFYEGTFN-WESVQKHDAASLLKLFIRELPOPLLSVEYL 375  
QY 190 RDAFRTL 196  
Db 376 K-AFAQV 381

## RESULT 15

T14280

RWL protein - mouse

C:Species: Mus musculus (house mouse)

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999

C:Accession: T14280

R:Wilkinson, R.; Fitter, S.; Tscharke, D.; Simmons, A.

submitted to the EMBL Data Library, April 1998

A:Reference number: Z17958

A:Accession: T14280

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1829 &lt;WIL&gt;

A:Cross-references: EMBL:AF060565; NID:g3091277; PID:g3091278; PIDN:AAC15232.1

A:Experimental source: strain BALB/c; brain

Query Match 7.1%; Score 77; DB 2; Length 1829;  
Best Local Similarity 27.2%; Pred. No. 40;  
Matches 28; Conservative 15; Mismatches 26; Indels 34; Gaps 6;

QY 69 DLDPGRCYRVTF--WFTSWSPCYDCARHVADFLRGNNPLSLRIETARLYFCEDRKAEPEGL 126  
Db 678 DLEPGKSKIANIYFDPGLQCGD-HRYI-----GLPFUS-----KSEPK-- 715  
QY 127 RRLHRAGVQIAIMTFKDYFYCWTFVFNHRTFKAWGLHNS 169  
Db 716 -----VQPGVAMQED---LWNADWDAAHQSLFKAWMGIKENA 748

Search completed: June 14, 2003, 18:11:56

Job time : 44 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2003, 17:59:53 ; Search time 24 Seconds  
(without alignments)  
342.180 Million cell updates/sec

Title: US-09-966-880A-8

Perfect score: 1086

Sequence: 1 MDSLMMNRKFLYQKNRW.....ILLPLYEVDDLRAFRITGL 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	364	33.5	382	1 PHB3_HUMAN	Q9uh17 homo sapien
2	363.5	33.5	199	1 PHB1_HUMAN	P31941 homo sapien
3	339	31.2	190	1 PHB2_HUMAN	Q9ue74 homo sapien
4	225	20.7	236	1 ABME_RABIT	P47855 oryctolaqus
5	216.5	19.9	236	1 ABME_HUMAN	P41238 homo sapien
6	211	19.4	229	1 ABME_MOUSE	P51908 mus musculus
7	198	18.2	229	1 ABME_RAT	P38483 rattus norv
8	97.5	9.0	369	1 CT19_YEAST	Q02732 saccharomyc
9	85	7.8	503	1 MATK_KUNBA	Q9tkc0 kunzea baxt
10	78.5	7.2	817	1 YG4B_YEAST	P46951 saccharomyc
11	77.5	7.1	402	1 VGLD_PPRVI	P07645 pseudorabie
12	77	7.1	1829	1 RWL_MOUSE	O70472 mus musculus
13	75.5	7.0	382	1 HUGA_APIME	Q08169 apis mellif
14	75	6.9	354	1 WN14_CHICK	O42280 gallus gall
15	73.5	6.8	337	1 HNR_ECOLI	P37055 escherichia
16	73.5	6.8	482	1 6PGD_SHEEP	P00349 ovis aries
17	73.5	6.8	738	1 TRFM_MOUSE	Q9r0r1 mus musculus
18	72.5	6.7	191	1 INF_ANAPL	P51526 anas platyr
19	72.5	6.7	641	1 PAPB_MOUSE	Q9wvp6 mus musculus
20	72	6.6	357	1 WN15_HUMAN	O14905 homo sapien
21	72	6.6	433	1 ERF1_SCHPO	P79063 schizosacch
22	72	6.6	792	1 GUAA_CAEEL	Q09580 caenorhabdi
23	72	6.6	811	1 SYM_TREPA	O83776 treponema p
24	72	6.6	1094	1 DRQL_GPCMV	Q69025 guinea pig
25	72	6.6	3866	1 HRX_MOUSE	P55200 mus musculus
26	72	6.6	3969	1 HRX_HUMAN	Q03164 homo sapien
27	71.5	6.6	505	1 PHTL_BRAJA	P23222 bradyrhizob
28	71	6.5	330	1 FXTD_COMTE	Q59727 comamonas t
29	71	6.5	718	1 PERM_MOUSE	P11247 mus musculus
30	71	6.5	925	1 DBL_HUMAN	P10911 homo sapien
31	71	6.5	1805	1 RWI_HUMAN	Q92545 homo sapien
32	70.5	6.5	630	1 F1MB_CHICK	P19179 gallus gall
33	70	6.4	626	1 HCY6_ANDAU	P80476 androctonus

34	70	6.4	1290	1 BXB_CLOBO	P10844 clostridium
35	69.5	6.4	493	1 YMDC_ECOLI	P75919 escherichia
36	69.5	6.4	636	1 PAPB_HUMAN	Q9nrj5 homo sapien
37	69	6.4	900	1 GLND_PSEAE	Q9z9h0 pseudomonas
38	69	6.4	1070	1 EX5C_BUCAI	P57528 buchnera ap
39	68.5	6.3	528	1 HEXA_MOUSE	P29416 mus musculus
40	68.5	6.3	682	1 KIF2_XENLA	Q91637 xenopus lae
41	68.5	6.3	988	1 TNP9_ECOLI	P51565 escherichia
42	68.5	6.3	996	1 VGNM_KCMV	P13561 red clover
43	68.5	6.3	3414	1 POLG_LANVT	P29837 l genome po
44	68	6.3	332	1 EXPA_CEPAC	P11935 c cephalosp
45	68	6.3	387	1 Y4PF_RHISN	P55615 rhizobium s

#### ALIGNMENTS

##### RESULT 1

ID	PHB3_HUMAN	STANDARD;	PRT;	382 AA.
AC	Q9UH17: Q95618;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Phorbolin 3 (APOBEC1-like).			
GN	APOBEC1L.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20057165; PubMed=10591208;			
RA	Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,			
RA	Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,			
RA	Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,			
RA	Bird C.P., Blakey S.E., Bridgman A.M., Buck D., Burgess J.,			
RA	Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,			
RA	Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,			
RA	Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,			
RA	Dhani P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,			
RA	Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,			
RA	Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,			
RA	Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,			
RA	Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,			
RA	Laird G.K., Langford C.F., Levensha M.A., Lloyd C., Lloyd D.M.,			
RA	Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,			
RA	Mcclay J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,			
RA	Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,			
RA	Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,			
RA	Scott C.E., Sehra H.K., Skuce C.D., Smalley S., Smith M.L.,			
RA	Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,			
RA	Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,			
RA	Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,			
RA	Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,			
RA	Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,			
RA	Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyma S.,			
RA	Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,			
RA	Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,			
RA	Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,			
RA	Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,			
RA	Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,			
RA	Zhan M., Zhang G., Chisoe S., Murray J., Miller N., Minx P.,			
RA	Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,			
RA	Cordes M., Du K., Fulton L., Goela D., Graves T., Hawkins J.,			
RA	Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,			
RA	Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,			
RA	Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,			
RA	Emanuel B.S., Shaikh T., Kurahashi H., Saitta S., Budarf M.L.,			
RA	Mcdermid H.E., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,			
RA	Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,			
RA	Seroussi E., Fransson I., Tapia I., Bruder C.E., O'Brien K.P.,			
RA	Wilkinson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,			

RA Ttlahun Y., Wright H.;  
 RT "The DNA sequence of human chromosome 22.";  
 RL Nature 402:489-495(1999).  
 CC [2]  
 RP SEQUENCE OF 148-382 FROM N.A.  
 RC TISSUE-Keratinocytes;  
 RA Madsen P.;  
 RT "Molecular cloning of phorbol 1.3";  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES  
 CC FAMILY. STRONG, TO APOLIPROTEIN B MRNA EDITING PROTEIN.  
 CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-148 IS THE  
 CC INITIATOR.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: AL022318; CAB45270.1; -;  
 DR EMBL; U61084; AAD00090.1; -;  
 DR InterPro: IPR002125; dCMP/cyt\_deam.  
 DR PROSITE: PS00903; CYT\_DCMP\_DEAMINASES; 2.  
 KW Hydrolase.  
 SQ SEQUENCE 382 AA; 45924 MW; DA6EDD23E8856240 CRC64;

Query Match 33.5%; Score 364; DB 1; Length 382;  
 Best Local Similarity 43.5%; Pred. No. 2.5e-29;  
 Matches 83; Conservative 26; Mismatches 62; Indels 20; Gaps 6;  
 QY 5 LMRKRLYQFKNVWAKGRRETYLCYVVKRRDSATSFSLD--FGYLRN--KN-----G 54  
 Db 192 LMDPDTFTFNNDPLVLRRTTYLCYEVERLDNGTWLMDQHWGFLCNKLLCGFYG 251  
 QY 55 CHVELLFLRYISDWLDPGRCYRVTFWTSWSPCYD--CARHVADEFLRGPNLSLRIFTAR 112  
 Db 252 RHAEFLRLDVLPSLQDPAQIYRVTFWTSWSPCFSGCAGEVRAFLQENTHVRILFAAR 311  
 QY 113 LYFCEDRKAP---EGLRLHRAVQVIAIMTFKDYFCWNTFVENHRTFKAWGLHENS 169  
 Db 312 IY-----DYDPLYKEALQMLRDAGAQSIMTYDEFCYDWFYRGCPFPQWDGLEHS 366  
 QY 170 VRLSRLRL 180  
 Db 367 QALSGRLRAIL 377

RESULT 2  
 PHB1\_HUMAN STANDARD; PRT; 199 AA.  
 ID PHB1\_HUMAN  
 AC P31941; Q12807;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Phorbol 1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Epidermis;  
 RX MEDLINE=9939284; PubMed=10469298;  
 RA Madsen P.P., Anant S., Rasmussen H.H., Gromov P., Vorum H.,  
 RA Dumanski J.P., Tommerup N., Collins J.E., Wright C.L., Dunham I.,  
 RA Macginnitie A.J., Davidson N.O., Celis J.E.;  
 RT "psoriasis up-regulated phorbol-1 shares structural but not  
 RT functional similarity to the mRNA-editing protein apobec-1";  
 RL J. Invest. Dermatol. 113:162-169(1999).  
 CC [2]

RP SEQUENCE OF 53-60; 112-121 AND 129-137.  
 RC TISSUE-Keratinocytes;  
 RX MEDLINE=93162043; PubMed=1286667;  
 RA Rasmussen H.H., van Damme J., Puype M., Gesser B., Celis J.E.,  
 RA Vandekerckhove J.;  
 RT "Microsequences of 145 proteins recorded in the two-dimensional gel  
 RT protein database of normal human epidermal keratinocytes.";  
 RL Electrophoresis 13:960-969(1992).  
 CC -!- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES  
 CC FAMILY. STRONG, TO APOLIPROTEIN B MRNA EDITING PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: U03891; AAA03706.2; -;  
 DR Aarhus/Ghent-2DPAGE; 2116; IEF.  
 DR InterPro: IPR002125; dCMP/cyt\_deam.  
 DR PROSITE: PS00903; CYT\_DCMP\_DEAMINASES; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 199 AA; 23012 MW; 42E99E0D7DF7AA14 CRC64;

Query Match 33.5%; Score 363.5; DB 1; Length 199;  
 Best Local Similarity 43.5%; Pred. No. 1.3e-29;  
 Matches 83; Conservative 28; Mismatches 57; Indels 23; Gaps 7;  
 QY 5 LMRKRLYQFKNVWAKGRRETYLCYVVKRRDSATSFSLD--FGYLRN--KN-----G 54  
 Db 12 LMDPDTFTFNNDPLVLRRTTYLCYEVERLDNGTWLMDQHWGFLCNKLLCGFYG 68  
 QY 55 CHVELLFLRYISDWLDPGRCYRVTFWTSWSPCYD--CARHVADEFLRGPNLSLRIFTAR 112  
 Db 69 RHAEFLRLDVLPSLQDPAQIYRVTFWTSWSPCFSGCAGEVRAFLQENTHVRILFAAR 128  
 QY 113 LYFCEDRKAP---EGLRLHRAVQVIAIMTFKDYFCWNTFVENHRTFKAWGLHENS 169  
 Db 129 IY-----DYDPLYKEALQMLRDAGAQSIMTYDEFCYDWFYRGCPFPQWDGLEHS 183  
 QY 170 VRLSRLRL 180  
 Db 184 QALSGRLRAIL 194

RESULT 3  
 PHB2\_HUMAN STANDARD; PRT; 190 AA.  
 ID PHB2\_HUMAN  
 AC Q9UE74;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Phorbol 2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Keratinocytes;  
 RA Madsen P.;  
 RT "Molecular cloning of phorbol 2.";  
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.  
 CC -!- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES  
 CC FAMILY. STRONG, TO APOLIPROTEIN B MRNA EDITING PROTEIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial



CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; U61083; AAD00089.1; --  
DR IncePro; IPR002125; dCMP\_cyt\_deam.  
DR PROSITE; PS00903; CYT\_DCMP\_DEAMINASES; FALSE\_NEG.  
KW Hydrolase.  
SQ SEQUENCE 190 AA; 22453 MW; A54DCBC100FC26F3 CRC64;

Query Match 31.28; Score 339; DB 1; Length 190;  
Best Local Similarity 41.18; Pred. No. 3.7e-27;  
Matches 78; Conservative 27; Mismatches 65; Indels 20; Gaps 6;

QY 6 MNRKFLYQFNWAKGRRTYLYCYVVKRDSATSFSLD--FGYLRN--KN-----GC 55  
DB 1 MDPPTFFNFNDPLVLRRTGYLYCYEVEERLDNGWVLLDQHMGLFNCNEAKNLLCGFYGR 60  
QY 56 HVELLFYRISDMDLDFGRCYRVTVFTSWSPCYD--CARHVADEFLRGNPLSLRIFATRL 113  
DB 61 HADWRFDLVPSLQDPAQYRVTVFTSWSPCFESWGCAGEVRAFLQENTHVRPLPIFAARI 120  
QY 114 YFCEDRAEP---EGLRLRHAGVQIAIMTFKDYFCWNTVEVHERTFKAWEGHNSV 170  
DB 121 Y-----DYDPLYKEALQMLRDAGAQSIMTYDEFCYWDTEVYRQCGFPQPDWGLEHSQ 175  
QY 171 RLSRLRLRL 180  
DB 176 ALSRLRL 185

RESULT 4  
ABME\_RABBIT  
ID ABME\_RABBIT STANDARD; PRT; 236 AA.  
AC P47855;  
DT 01-FEB-1996 (Rel. 33, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Apolipoprotein B mRNA editing protein (APOBEC-1).  
GN APOBEC1.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
RC STRAIN-New Zealand white; TISSUE=Small intestine;  
RX MEDLINE=94342367; PubMed=8063816;  
RA Yamanaka S., Poksay K.S., Balestra M.E., Zeng G.-Q., Innerarity T.L.;  
RT "Cloning and mutagenesis of the rabbit Apob mRNA editing protein. A  
RT zinc motif is essential for catalytic activity, and noncatalytic  
RT auxiliary factor(s) of the editing complex are widely distributed.";  
RL J. Biol. Chem. 269:21725-21734(1994).  
CC -!- FUNCTION: RESPONSIBLE FOR THE POSTTRANSCRIPTIONAL EDITING OF  
CC A CAA CODON FOR GLN TO A UAA CODON FOR STOP IN THE APOB MRNA.  
CC -!- COFACTOR: ZINC.  
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN THE INTESTINE.  
CC -!- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES  
CC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
DR EMBL; U10695; AAA56718.1; --  
DR InterPro; IPR002125; dCMP\_cyt\_deam.  
DR Pfam; PF00383; dCMP\_cyt\_deam; 1.  
DR PROSITE; PS00903; CYT\_DCMP\_DEAMINASES; 1.  
KW mRNA processing; Hydrolase; Zinc.

FT METAL 61 ZINC (BY SIMILARITY).  
FT METAL 93 ZINC (BY SIMILARITY).  
FT METAL 96 ZINC (BY SIMILARITY).  
FT DOMAIN 180 LEU-RICH MOTIF.  
FT MUTAGEN 61 H->A: NONE OR LITTLE EDITING ACTIVITY.  
FT MUTAGEN 61 H->C: RETAINS MOST EDITING ACTIVITY.  
FT MUTAGEN 63 E->A: NONE OR LITTLE EDITING ACTIVITY.  
FT MUTAGEN 92 P->A: RETAINS MOST EDITING ACTIVITY.  
FT MUTAGEN 93 C->A: NONE OR LITTLE EDITING ACTIVITY.  
FT MUTAGEN 96 C->A: NONE OR LITTLE EDITING ACTIVITY.  
SQ SEQUENCE 236 AA; 27719 MW; AB3041CA5102F13 CRC64;

Query Match 20.7%; Score 225; DB 1; Length 236;  
Best Local Similarity 37.4%; Pred. No. 1.5e-15;  
Matches 49; Conservative 26; Mismatches 48; Indels 8; Gaps 4;  
QY 24 RRETYLYCYVVKRRDSATSFSLDFGYLRNKGCG-HVELLFL-RYISDMDLDFGRCYRVTVWF 81  
DB 33 RKEACLLYEIKWGASSKTRSS-----GKNTTNHVEVNFLEKLTSEGLGSPSTCCSITWF 87  
QY 82 TSWSPCYDCARHVADEFLRGNPLSLRIFATRYFCEDRAEPGLRLRHAGVQIAIMTF 141  
DB 88 LSWSPCFESWMAIREFLSQHGPVTLIFVARLFQHMRRR-NRQGLKDLVTSGTVKRVMSV 146  
QY 142 KDIFYCWNTFV 152  
DB 147 SEYCYCWENFV 157

RESULT 5  
ABME\_HUMAN  
ID ABME\_HUMAN STANDARD; PRT; 236 AA.  
AC P41238;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1996 (Rel. 33, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Apolipoprotein B mRNA editing protein (APOBEC-1) (HEPR).  
GN APOBEC1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Intestine;  
RX MEDLINE=94268910; PubMed=8208612;  
RA Hadjiagapiou C., Giannoni F., Funahashi T., Skarosi S.F.,  
RT Davidson N.O.;  
RT "Molecular cloning of a human small intestinal apolipoprotein B mRNA  
RT editing protein.";  
RL Nucleic Acids Res. 22:1874-1879(1994).  
RN [2]  
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.  
RC TISSUE=Small intestine;  
RX MEDLINE=94359963; PubMed=8078915;  
RA Lau P., Zhu H.-J., Baldini A., Charnsangavej C., Chan L.;  
RT "Dimeric structure of a human apolipoprotein B mRNA editing protein  
RT and cloning and chromosomal localization of its gene.";  
RL Proc. Natl. Acad. Sci. U.S.A. 91:8522-8526(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Peripheral blood leukocytes;  
RX MEDLINE=98140126; PubMed=9479499;  
RA Fujino T., Navaratnam N., Scott J.;  
RT "Human apolipoprotein B RNA editing deaminase gene (APOBEC1).";  
RL Genomics 47:266-275(1998).  
CC -!- FUNCTION: RESPONSIBLE FOR THE POSTTRANSCRIPTIONAL EDITING OF  
CC A CAA CODON FOR GLN TO A UAA CODON FOR STOP IN THE APOB MRNA.  
CC -!- COFACTOR: ZINC (BY SIMILARITY).  
CC -!- SUBUNIT: HOMODIMER.  
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN THE SMALL INTESTINE.  
CC -!- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES  
CC FAMILY.



```

CC -!- FUNCTION: RESPONSIBLE FOR THE POSTTRANSCRIPTIONAL EDITING OF
CC A CAA CODON FOR GLN TO A UAA CODON FOR STOP IN THE APOB MRNA.
CC -!- COFACTOR: ZINC (BY SIMILARITY).
CC -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN THE LIVER AS WELL AS SMALL
CC INTESTINE.
CC -!- SIMILARITY: BELONGS TO THE CYTIDINE AND DEOXYCYTIDYLATE DEAMINASES
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; L07114; AAA17394.1; -;
DR InterPro; IPR002125; dCMP/cyt_deam.
DR Pfam; PF00383; dCMP_cyt_deam; 1.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; 1.
KW MRNA processing; Hydrolase; Zinc.
FT METAL 61 61 ZINC (BY SIMILARITY).
FT METAL 93 93 ZINC (BY SIMILARITY).
FT METAL 96 96 ZINC (BY SIMILARITY).
FT METAL 180 193 LEU-RICH MOTIF.
SQ DOMAIN
SQ SEQUENCE 229 AA; 27274 MW; 08766441882789B3 CRC64;

Query Match 18.2%; Score 198; DB 1; Length 229;
Best Local Similarity 32.9%; Pred. No. 7.6e-13;
Matches 56; Conservative 24; Mismatches 62; Indels 28; Gaps 6;

Qy 24 RRETYLCVVK-----RRDSATSFSLDFGLRNKNGCHVELFL-RYISDWLDLPGR 74
Db 33 RKETCLLYEINWGGHSHWRTSNT-----NK---HVENVFIEKFTERYFCENT 80
Qy 75 CYRVTFWTSNCPYCDCAHVADFLRGPNLSLRIFTARLYPCEDKAEPEGLRLRHAGV 134
Db 81 RCSITWFLSPWCGECSRAITFEFLSRYPHTVLTFTYIARLYHADPR-NRQGLRDLISGV 139
Qy 135 QIAIWTFKDYCWNTFFVENERTFKAWEGHLSNVRSLRQLRLILPLY 184
Db 140 TIQIMTEQESGYCWRNFVNSPFSNEAHWPYPHVLVRL-----YVLELY 183

RESULT 8
CT19_YEAST
ID CT19_YEAST STANDARD; PRT; 369 AA.
AC Q02732;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Kinetochores protein CT19.
GN CT19 OR IPLO18W OR LPB13W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Hyland K.M., Hieter P.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,
RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster E., Davis K., Davis R.W.,
RA Dierich F.S., Delius H., Dipolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,

```

```

RA Hunnicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharfe M.,
RA Scherens B., Schramm S., Schroeder M., Solcu A.M., Tettelin H.,
RA Urrestazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL Nature 387:103-105(1997).
CC -!- FUNCTION: IMPORTANT FOR CHROMOSOME SEGREGATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; U72265; AAB17275.1; -;
DR EMBL; U36624; AAB68169.1; -;
DR SGD; S0005939; CTF19.
SQ SEQUENCE 369 AA; 42782 MW; 23B4CBD6AE26E793 CRC64;

Query Match 9.0%; Score 97.5; DB 1; Length 369;
Best Local Similarity 22.2%; Pred. No. 0.018;
Matches 50; Conservative 34; Mismatches 70; Indels 71; Gaps 11;

Qy 2 DSLMNRKFLYOFKNVRWAKGRRETYLCYVYVKKRDSATSFSLDFGLRNKNGCHVELL- 60
Db 62 DALLFRNTLLQEI-----QTYQNILMKENSKT-----KNG---DILQ 97
Qy 61 -----FLYISDWLDPG-----RCYRVTFWTSN-----PCYDCARHVADFLRGPN 103
Db 98 NDIQDFNLNLSISSNPNSAISDRKRVRLNGLTNLQELVTKYDTPLL-----N 149
Qy 104 LSLRTFTARLYPCEDKAEPEGLRLRHAGVQIAITWTFKDYFCWNT-----FVE 153
Db 150 NMLRLSYLDRHYPHLVQSVQSRDRVHNDGIEVLVYNYK---FCRTMPPFEIQKMFYK 206
Qy 154 NHRTFKAWEGHL-ENSVRL-----SQLRRILLPLYEVDLR 190
Db 207 PEDSTLLKWEILRISTNVRLKAKQLLATRNFKCLLSLYEFDKIK 251

RESULT 9
MATK_KUNBA
ID MATK_KUNBA STANDARD; PRT; 503 AA.
AC Q9TKC0;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable intron maturase (Maturase K).
GN MATK.
OS Kunzea baxteri (Scarlet kunzea).
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Myrtales; Myrtaceae; Kunzea.
OX NCBI_TaxID=106042;
RN [1]
RP SEQUENCE FROM N.A.
RA O'Brien M.M., Wilson P.G., Quinn C.J.;
RT "Molecular systematics of the Leptospermum suballiance (Myrtaceae).";
RL Aust. J. Bot. 48:621-628(2000).
CC -!- FUNCTION: PROBABLY ASSISTS IN SPLICING CHLOROPLAST GROUP II
CC INTRONS.
CC -!- SIMILARITY: WITH CORRESPONDING ORF IN OTHER PLANT CHLOROPLASTS,
CC AND REGIONS OF SIMILARITY TO MATURASE-LIKE POLYPEPTIDES ENCODED
CC BY MITOCHONDRIAL INTRONS.
CC -----

```

Query Match 7.28; Score 78.5; DB 1; Length 817;

[illegible]

```

RC TISSUE=Venom;
RA Jacobson R.S., Hoffman D.R., Kemeny D.M.;
RT "The cross-reactivity between bee and vespid hyaluronidases has a structural basis.";
RL J. Allergy Clin. Immunol. 89:292-292(1991).
RN [3]
RP CARBOHYDRATE-LINKAGE SITE ASN-263.
RX MEDLINE=20455243; PubMed=10998264;
RA Kolarich D., Altmann F.;
RT "N-glycan analysis by matrix-assisted laser desorption/ionization mass spectrometry of electrophoretically separated nonmammalian proteins: application to peanut allergen Ara h 1 and olive pollen allergen Ole e 1.";
RL Anal. Biochem. 285:64-75(2000).
RR -|- FUNCTION: MAY PLAY A ROLE IN REPRODUCTION.
CC -|- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-linkages between N-acetyl-beta-D-glucosamine and D-glucuronic acid residues in hyaluronate.
CC -|- TISSUE SPECIFICITY: EXPRESSED IN THE VENOM GLANDS OF WORKER BEES. IT IS ALSO DETECTED IN THE TESTES OF DRONES BUT NOT IN THE QUEEN-BEE VENOM GLANDS OR IN PUPAE.
CC -|- PTM: TWO DISULFIDE BONDS MAY BE PRESENT.
CC -|- SIMILARITY: BELONGS TO FAMILY 56 OF GLYCOSYL HYDROLASES.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement or send an email to licensee@isb-sib.ch).
CC -----
DD EMBL; LI0710; AAA27730.1; -.
DR PIR; A47477; A47477.
DR GlycoSuiteDB; Q08169; -.
DR InterPro; IPR001968; GH_56.
DR Pfam; PF01630; Glyco_hydro_56; 1.
DR PRINTS; PR00846; GLYHDLRASE5.
DR PRODOM; PD003549; GH_56; 1.
KW Hydrolase; Glycosidase; Signal; Glycoprotein; Zymogen; Venom;
KX Allergen.
FT SIGNAL          1      28       OR 24 (POTENTIAL).
FT PROPSEP        29      33         POTENTIAL.
FT CHAIN           34     382        HYALURONOGUCOSAMINIDASE.
FT CARBOHYD        115    115        N-LINKED (GLCNAC..) (POTENTIAL)..
FT CARBOHYD        263    263        N-LINKED (GLCNAC..) (COMPLEX)..
FT VARIANT         371    371        D -> S (IN CLONE HYA-2)..
FT CONFLICT        37      37        N > D (IN REF..2)..
SQ SEQUENCE        382 AA; 44259 MW; 3E68Z22E9SCA1B56 CRC64;
Query Match              7.0%; Score 75.5; DB 1; Length 382;
Best Local Similarity   21.6%; Pred.No. 3.1;
Matches                 36; Conservative 21; Mismatches 51; Indels 59; Gaps 7;
QY 81 FTSWSPCYDCARHVDFLRGNPNLSRIFFARTFLRYFCEDRKAAPEGLRLRHAGVQAIAIMT 140
Db 144 FEWSMRPIF---RQNWASLOPYKKLSVEVRREHPFDWDQRVQEAKRRFEKYGOLFMEET 200
QY 141 FK-----DYFFCWN-----TFVENHERT-----FK 160
Db 201 LKAARKRMPPAANGYYAYPCYNLTNPQSAOCEATTMQNDKMKSWLFESEVDLLPSVYL 260
QY 161 AWB-----GLHENSVRLSRQL---RRILLPL--YEVDLDLR 191
Db 261 RWNLTSGERVLGVGRVKEARLTRARMTTSRKKVLPYYNYKDORRD 307

RESULT 14
WN14_CHUNK
AD WN14_CHUNK STANDARD; PRT; 354 AA.
IC O42280;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
```



```
-----NOT ANNOTATED_CDS.
DR EMBL; M64675; ; NOT ANNOTATED_CDS.
DR EMBL; X66003; CAA46802.1; -.
DR EMBL; AE000222; AAC74317.1; -.
DR EMBL; D90758; BAA36103.1; -.
DR EMBL; D90759; BAA36115.1; -.
DR EMBL; D90852; BAA16029.1; -.
DR EMBL; AE005341; AAG56092.1; -.
DR EMBL; AP002556; BAB35160.1; -.
DR PIR; S28505; S28505
DR PIR; A36871; A36871.
DR HSP; P52934; IQMP.
DR EcoGene; EG12121; hnr.
DR InterPro; IPR001789; Response_reg.
DR Pfam; PF00072; response_reg.1.
DR ProDom; PD000039; Response_reg; 1.
DR SMART; SM00448; REC.1.
DR PROSITE; PS0110; RESPONSE_REGULATORY; 1.
KW Phosphorylation; Transcription regulation; DNA-binding;
KW Sensory transduction; Complete proteome.
FT DOMAIN 9 123 RESPONSE_REGULATORY.
FT MOD_RES 58 58 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 337 AA; 37302 MW; AB962EF94BC7B470 CRC64;

Query Match 6.8%; Score 73.5; DB 1; Length 337;
Best Local Similarity 21.6%; Pred. No. 4.2;
Matches 32; Conservative 19; Mismatches 52; Indels 45; Gaps 5;

Qy 79 TWFTSWSPCYDCARHVADFLRGNPMLSRIFTARLYFCEDRKAEPGLRLH----- 130
Db :||:| | | | | | | | | | | | | | | | | | | | | | | | | | | |
25 SWFSSLGATTVLAADGVDALE-----LLGGFTPLMICDIAMPRMNGKLLLEHNRGDQ 79
Qy 131 -----RAGVO-IAIMTFKD-----YFYC-----WNTFVENHRT 158
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
80 TPVLVISATENMADIKALRLGVEDVLKPKVNDLNRLEMTVFACLYPSMFNSRVEEERL 139

Qy 159 FKAMEGLHENSRLSRQLRILLPLYEV 186
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
140 FRDWDAMVDNPAAAAKLLQELQPPVQV 167
```

Search completed: June 14, 2003, 18:09:42  
Job time : 26 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2003, 18:05:13 ; Search time 78 Seconds  
(without alignments)  
523.043 Million cell updates/sec

Title: US-09-966-880A-8  
Perfect score: 1086  
Sequence: 1 MDSLIMNRKFLYQKQVNRW.....ILLPLYEVDLDRFAFTLGL 198

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_21.\*  
1: sp\_archaea.\*  
2: sp\_bacteria.\*  
3: sp\_fungi.\*  
4: sp\_human.\*  
5: sp\_invertebrate.\*  
6: sp\_mammal.\*  
7: sp\_mhc.\*  
8: sp\_organelle.\*  
9: sp\_phase.\*  
10: sp\_plant.\*  
11: sp\_rodent.\*  
12: sp\_virus.\*  
13: sp\_vertebrate.\*  
14: sp\_unclassified.\*  
15: sp\_virus.\*  
16: sp\_bacteriaph.\*  
17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1086	100.0	198	4	Q9GZX7
2	1008	92.8	198	11	Q9WVE0
3	390	35.9	384	4	Q9HCL6
4	369.5	34.0	204	4	Q9Y555
5	367.5	33.8	190	4	Q96F12
6	354	32.6	294	4	Q9NRW3
7	349	32.1	429	11	Q99J72
8	336.5	31.0	386	4	Q96AK3
9	315.5	29.1	139	4	Q9Y4V1
10	278	25.6	224	11	Q9WV35
11	274	25.2	224	4	Q9Y235
12	270.5	24.9	121	4	Q9Y553
13	257.5	23.7	103	4	Q9Y550
14	251	23.1	111	4	Q9UH18
15	240.5	22.1	121	4	Q9Y554
16	238.5	22.0	131	4	Q95930

17	238.5	22.0	245	4	Q9Y549	Q9Y549 homo sapien
18	219	20.2	229	11	Q9EQP0	Q9eqp0 mesocricetu
19	218.5	20.1	236	4	Q9UM71	Q9um71 homo sapien
20	213.5	19.7	235	6	Q9TUI7	Q9tui7 monodelphis
21	211	19.4	229	11	Q99L67	Q99l67 mus musculus
22	198	18.2	127	4	Q9H9H8	Q9h9h8 homo sapien
23	182.5	16.8	77	4	Q9Y551	Q9y551 homo sapien
24	179	16.5	132	4	Q9UE64	Q9ue64 homo sapien
25	102	9.4	720	5	Q9W293	Q9w293 drosophila
26	85	7.8	414	17	Q97L68	Q97l68 sulfolobus
27	84.5	7.8	504	8	Q9TKC5	Q9tkc5 astromyrtu
28	83.5	7.7	504	8	Q9TKC4	Q9tkc4 astromyrtu
29	83.5	7.7	533	11	O70276	O70276 rattus norv
30	82.5	7.6	483	11	Q8VHP2	Q8vhp2 mus musculu
31	82.5	7.6	533	11	Q91Z05	Q91z05 mus musculu
32	81.5	7.5	400	12	Q9IW02	Q9iw02 pseudorabie
33	81.5	7.5	516	12	Q9IW00	Q9iw00 pseudorabie
34	81.5	7.5	2342	5	O46048	O46048 drosophila
35	80.5	7.4	505	8	Q9TKB9	Q9tkb9 kunzea capi
36	80	7.4	476	16	O25366	O25366 helicobacte
37	79.5	7.3	503	8	Q9TKC3	Q9tkc3 astromyrtu
38	79.5	7.3	504	8	Q9MSA5	Q9msa5 podocarpus
39	78.5	7.2	503	8	Q9TKB1	Q9tkb1 leptospermu
40	78	7.2	504	8	Q95GT9	Q95gt9 nepenthes m
41	77.5	7.1	251	5	O17138	O17138 caenorhabdi
42	77.5	7.1	279	12	Q91F41	Q91f41 cydia pomon
43	77.5	7.1	369	5	Q22995	Q22995 caenorhabdi
44	77.5	7.1	404	12	Q92290	Q92290 pseudorabie
45	77.5	7.1	618	4	Q96S64	Q96s64 homo sapien

## ALIGNMENTS

RESULT 1

Q9GZX7	PRELIMINARY;	PRT;	198 AA.
ID	Q9GZX7		
AC	Q9GZX7;		
DT	01-MAR-2001 (TREMREL. 16, Created)		
DT	01-MAR-2001 (TREMREL. 16, Last sequence update)		
DT	01-MAR-2002 (TREMREL. 20, Last annotation update)		
DE	Activation-induced cytidine deaminase.		
GN	AID.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20408890; PubMed=10950930;		
RA	Muto T., Muranatsu M., Taniwaki M., Kinoshita K., Honjo T.;		
RT	"Isolation, tissue distribution and chromosomal localization of the		
RT	human activation-induced cytidine deaminase (hAID) gene."		
RL	Genomics 68:85-88(2000).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20460541; PubMed=11007475;		
RA	Revy P., Muto T., Levy Y., Geissmann F., Plebani A., Sanal O.,		
RA	Catalan N., Forveille M., Dufourcq-Lagelouse R., Gennery A.,		
RA	Tezcan I., Ersoy F., Kayserili H., Ugazio A.G., Brousse N.,		
RA	Muranatsu M., Notarangelo L.D., Kinoshita K., Honjo T., Fischer A.,		
RT	Durandy A.;		
RT	"Activation-induced cytidine deaminase (AID) deficiency causes the		
RT	autosomal recessive form of the Hyper-IgM syndrome (HIGM2).";		
RL	Cell 102:565-575(2000).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-B-CELL;		
RA	Strausberg R.;		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AB040431; BAB12721.1; -		
DR	EMBL; AB040430; BAB12720.1; -		
DR	EMBL; BC006296; AAH06296.1; -		

```
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 198 AA; 23953 MW; 3C27BB143DB184A9 CRC64;

Query Match 100.0%; Score 1086; DB 4; Length 198;
Best Local Similarity 100.0%; Pred. No. 7.4e-102; Indels 0; Gaps 0;
Matches 198; Conservative 0; Mismatches 0;

QY 1 MDSLLMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60
DB 1 MDSLLMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60

QY 61 FLRYISDMDLPGRCYRVYTWFTSWSPCYDCARHVADEFLRGNPNLSLRIFTARLYFCEDRK 120
DB 61 FLRYISDMDLPGRCYRVYTWFTSWSPCYDCARHVADEFLRGNPNLSLRIFTARLYFCEDRK 120

QY 121 APEGLRLRLHRAVGVOIAIMTFKDYFCYCNWTFVFNHERTFKAWEGHLHNSVRLSROLRRL 180
DB 121 APEGLRLRLHRAVGVOIAIMTFKDYFCYCNWTFVFNHERTFKAWEGHLHNSVRLSROLRRL 180

QY 181 LPLYEVDDLRLDAFRITGL 198
DB 181 LPLYEVDDLRLDAFRITGL 198

RESULT 2
Q9WVE0 PRELIMINARY; PRT; 198 AA.
AC Q9WVE0;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
DE Activation-induced cytidine deaminase.
GN AICDA OR AID.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=99303612; PubMed=10373455;
RA Muramatsu M., Sankaranand V.S., Anant S., Sugai M., Kinoshita K.,
RA Davidson N.O., Honjo T.;
RT "Specific expression of activation-induced cytidine deaminase (AID), a
RT novel member of the RNA-editing deaminase family in germinal center B
RT cells.";
RL J. Biol. Chem. 274:18470-18476(1999).
DR EMBL; AF132979; AAD41793.1; -.
DR MGD; MGI:1342279; Aicda.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 198 AA; 24030 MW; 18A3BA10CA54BEB2 CRC64;

Query Match 92.8%; Score 1008; DB 11; Length 198;
Best Local Similarity 92.9%; Pred. No. 5.5e-94;
Matches 183; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

QY 1 MDSLLMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60
DB 1 MDSLLMNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVELL 60

QY 61 FLRYISDMDLPGRCYRVYTWFTSWSPCYDCARHVADEFLRGNPNLSLRIFTARLYFCEDRK 120
DB 61 FLRYISDMDLPGRCYRVYTWFTSWSPCYDCARHVADEFLRGNPNLSLRIFTARLYFCEDRK 120

QY 121 APEGLRLRLHRAVGVOIAIMTFKDYFCYCNWTFVFNHERTFKAWEGHLHNSVRLSROLRRL 180
DB 121 APEGLRLRLHRAVGVOIAIMTFKDYFCYCNWTFVFNHERTFKAWEGHLHNSVRLSROLRRL 180

QY 181 LPLYEVDDLRLDAFRITGL 197
DB 181 LPLYEVDDLRLDAFRITGL 197

RESULT 3
Q9HC16 PRELIMINARY; PRT; 384 AA.
AC Q9HC16;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE MDS019 (Phorbol-like protein MDS019).
GN MDS019.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP Huang C., Qian B., Tu Y., Gu W., Wang Y., Han Z., Chen Z.;
RA "Novel genes expressed in hematopoietic stem/progenitor cells from
RA myelodysplastic Syndromes patient.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF182420; AAG14956.1; -.
DR EMBL; BC024268; AAH24268.1; -.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 384 AA; 46408 MW; 60525DC3B7D903D6 CRC64;

Query Match 35.9%; Score 390; DB 4; Length 384;
Best Local Similarity 44.9%; Pred. No. 3.1e-31;
Matches 83; Conservative 31; Mismatches 59; Indels 12; Gaps 3;

QY 6 MNRKFLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLD--FGYLRNK-----NGC 55
DB 197 MDPPTFTFNFNPNWVRGRHETLYCYEYERHNDTWLLNQRRGLCNQAPKHGFLRGR 256

QY 56 HVELLFLRYISDMDLPGRCYRVYTWFTSWSPCYDCARHVADEFLRGNPNLSLRIFTARLYF 115
DB 257 HAEICFLDVIPIFWKLLDDQDQIRVTCFTSWSPCFCAQEMAKFISKNKHVSLCIFTARIY- 315

QY 116 CEDRAKEPEGLRLHRAVGVOIAIMTFKDYFCYCNWTFVFNHERTFKAWEGHLHNSVRLSRQ 175
DB 316 --DQGRCEGLRTLAEGAKISIMTYSEFKHCWDTFVDHQCPCFPDGLDEHSQDLGR 374

QY 176 LRRIL 180
DB 375 LRAIL 379

RESULT 4
Q9Y555 PRELIMINARY; PRT; 204 AA.
AC Q9Y555;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE BK150C2.3 (Putative novel protein similar to APOBEC1 (Apolipoprotein B
DE mRNA editing protein) and Phorbol) (Fragment).
GN BK150C2.3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RP Matthews L.;
RA Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL022318; CAB45271.1; -.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
```

```
KW Lipoprotein.
FT NON_TER
SQ SEQUENCE 204 AA; 24285 MW; 79C656F580A40554 CRC64;

Query Match 34.0%; Score 369.5; DB 4; Length 204;
Best Local Similarity 44.9%; Pred. No. 1.8e-29;
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;

QY 11 FLYQFKNVWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----KNGCHVELLFLRY 64
DB 31 FYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKT--GVFRNQVDSETHCHAEKRCFLSW 88

QY 65 ISDWLDLPGRCYRVWTFTSWSPCYDCARHVADEFLRGNPNLSLRIFTARLYFCEDRKAEP 124
DB 89 FCDDILSPNTKYQVWTWTSWSPCPDCAGEVAEFLARHSNVNLTIFTARLYYFQ-YPCYQE 147

QY 125 GLRLRHAGVQIAIMTFKDYFCWNTFVENHERTFKAWEGHLSNVRSLRQLRILL 180
DB 148 GLRSLSQEGVAEIMDYEDFKYCWENFVYNDNEPFPKWKGLKTNFRLKRLRRESL 203

RESULT 5
Q96F12 PRELIMINARY; PRT; 190 AA.
AC Q96F12;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Similar to APOBEC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011739; AAH11739.1; -.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 190 AA; 22827 MW; DA0584EF75C91CF0 CRC64;

Query Match 33.8%; Score 367.5; DB 4; Length 190;
Best Local Similarity 44.9%; Pred. No. 2.6e-29;
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;

QY 11 FLYQFKNVWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----KNGCHVELLFLRY 64
DB 17 FYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKT--GVFRNQVDSETHCHAEKRCFLSW 74

QY 65 ISDWLDLPGRCYRVWTFTSWSPCYDCARHVADEFLRGNPNLSLRIFTARLYFCEDRKAEP 124
DB 75 FCDDILSPNTKYQVWTWTSWSPCPDCAGEVAEFLARHSNVNLTIFTARLYYFQ-YPCYQE 133

QY 125 GLRLRHAGVQIAIMTFKDYFCWNTFVENHERTFKAWEGHLSNVRSLRQLRILL 180
DB 134 GLRSLSQEGVAEIMDYEDFKYCWENFVYNDNEPFPKWKGLKTNFRLKRLRRESL 189

RESULT 6
Q9NRW3 PRELIMINARY; PRT; 294 AA.
AC Q9NRW3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Phorbolin I protein.
GN PBI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RA Gu J., Huang Q., Yu Y., Xu S., Wang Y., Han Z., Chen Z., Zhou J.,
RA Tu Y., Gu W., Fu G., Huang C.;
RT "Novel genes expressed in hematopoietic stem/progenitor cells from
RT Myelodysplastic Syndromes patient.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF165520; AAF86650.1; -.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
SQ SEQUENCE 294 AA; 33363 MW; IB39C7A13D690901 CRC64;

Query Match 32.6%; Score 354; DB 4; Length 294;
Best Local Similarity 43.9%; Pred. No. 9.8e-28;
Matches 75; Conservative 26; Mismatches 60; Indels 10; Gaps 5;

QY 11 FLYQFKNVWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----KNGCHVELLFLRY 64
DB 17 FYFQFKNLWEANDRNETWLCFTVEGIKRRSVVSWKT--GVFRNQVDSETHCHAEKRCFLSW 74

QY 65 ISDWLDLPGRCYRVWTFTSWSPCYDCARHVADEFLRGNPNLSLRIFTARLYFCEDRKAEP 124
DB 75 FCDDILSPNTKYQVWTWTSWSPCPDCAGEVAEFLARHSNVNLTIFTARLYYFQ-YPCYQE 133

QY 125 GLRLRHAGVQIAIMTFKDYFCWNTFVENHERTFKAWEGHLSNVRSLRQ 175
DB 134 GLRSLSQEGVAEIMDYEDFKYCWENFVYNDNEPFPKWKGI-KNOLSTSEK 183

RESULT 7
Q99J72 PRELIMINARY; PRT; 429 AA.
AC Q99J72;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Hypothetical 51.0 kDa protein.
GN BC003314.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003314; AAH03314.1; -.
DR MGD; MGI:193311; BC003314.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_2.
KW Hypothetical protein.
SQ SEQUENCE 429 AA; 51017 MW; BE44D01380AD7F6E CRC64;

Query Match 32.1%; Score 349; DB 11; Length 429;
Best Local Similarity 38.8%; Pred. No. 4.8e-27;
Matches 71; Conservative 36; Mismatches 72; Indels 4; Gaps 2;

QY 5 LNMRRKFLYQFKNVWAKGRRETYLCYVVKRDSATSFSLDFGYLRNKNCHVELLFLRY 64
DB 20 LISQETFPKFLNRYAIDRKDTFLCYEYVTRKDCSPVSLHGVFKKNINHAICFLYW 79

QY 65 ISDWLDLPGRCYRVWTFTSWSPCYDCARHVADEFLRGNPNLSLRIFTARLYFCEDRKA 121
DB 80 FHDKVLKVLSPREEFKITWYMSWSPCFECAEQVLRFLATHNLSLDFSSRLYNIRDEN 139

QY 122 EPEGLRRLRHAGVQIAIMTFKDYFCWNTFVENHERTFKAWEGHLSNVRSLRILL 181
DB 140 Q-QNLCLRLVQEGAQVAAMDLYEFKCKWKKEVDNGRRFRPWKLLTNFRYQDSKLOEILR 198

QY 182 PLY 184
DB 199 PCY 201
```

```
RESULT 8
Q96AK3
ID Q96AK3 PRELIMINARY; PRT; 386 AA.
AC Q96AK3
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE Hypothetical 46.6 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017022; AAH17022.1; -.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_2.
KW Hypothetical protein.
SQ SEQUENCE 386 AA; 46598 MW; 94C7253BDC85B22 CRC64;

Query Match 31.0%; Score 336.5; DB 4; Length 386;
Best Local Similarity 40.8%; Pred. No. 7.8e-26;
Matches 71; Conservative 27; Mismatches 71; Indels 5; Gaps 2;

QY 11 FLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRN----KNGCHVELLFLRYS 66
DB 213 FYFHKNLKACGRNESWLCFTMEVTKHHSVFRKRGVFRNOVDPTHCHAEKFCLSWFC 272
QY 67 DWLDLPGRCYRVWFTSPSCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPGL 126
DB 273 DDILSPNTNVEVWTSNPSCECAEVAEFLARHSNVNLITFTARLCYFWDYDQ-EG 331
QY 127 RRLHRAGVQIAIMTFKDYFYCWNTFVENHETFFKAWEGHNSVRLSRLRRL 180
DB 332 CSLSQEGASVKIMGYKDFVSCWKNFVSDDEPFKPMKGLQTNFLLKRLREIL 385

RESULT 9
Q94V1
ID Q94V1 PRELIMINARY; PRT; 139 AA.
AC Q94V1
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-2002 (TRENBLrel. 20, Last annotation update)
DE DJ494G10.1 (Novel protein similar to Phorbolin 1 and APOBEC1
DE (Apolipoprotein B mRNA editing enzyme, catalytic polypeptide 1))
DE (Fragment).
GN DJ494G10.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ramsay H.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL078641; CAB45659.1; -.
DR InterPro; IPR002125; dCMP/cyt_deam.
DR PROSITE; PS00903; CYT_DCMP_DEAMINASES; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 139 AA; 16075 MW; C77CB711DDAAA9C0 CRC64;

Query Match 29.1%; Score 315.5; DB 4; Length 139;
Best Local Similarity 48.5%; Pred. No. 3.2e-24;
Matches 65; Conservative 24; Mismatches 40; Indels 5; Gaps 2;

QY 47 GYLKNGKCHVELLFLRYSWDLDPGRCYRVWFTSPSCYDCARHVADFLRGNPNLSL 106
DB 6 GFL---EGRHAEELCFDVIPIFWKLDLDQDYRVCTFTSPSCYDCARHVADFLRGNPNLSL 62

RESULT 10
Q9WV35
ID Q9WV35 PRELIMINARY; PRT; 224 AA.
AC Q9WV35
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE APOBEC-2 protein.
GN APOBEC2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RX MEDLINE=99333690; PubMed=10403781;
RA Liao W., Hong S.H., Chan B.H., Rudolph F.B., Clark S.C., Chan L.;
RT "APOBEC-2, a cardiac- and skeletal muscle-specific member of the
RT cytidine deaminase supergene family.";
RL Biochem. Biophys. Res. Commun. 260:398-404(1999).
DR EMBL; AF161699; AAD45361.1; -.
DR MGD; MGI:1343178; Apobec2.
SQ SEQUENCE 224 AA; 25660 MW; 75F98BC2CF2EBE0A CRC64;

Query Match 25.6%; Score 278; DB 11; Length 224;
Best Local Similarity 36.7%; Pred. No. 3.4e-20;
Matches 65; Conservative 26; Mismatches 74; Indels 12; Gaps 5;

QY 11 FLYQFNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKN-GCHVELLFLRYSWD 69
DB 52 FKFOFRNVEYSSGRNKTFLCYVVEQSGGQAATQGYLEDEHAGAHAEAFNTILP-A 110
QY 70 LDPGRCYRVWFTSPSCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEP-GL 126
DB 111 FDPALKYNTVWVSSSPCAACADRLTKLTKNLLILVSLRFLMWE----EPEVQAAL 166
QY 127 RRLHRAGVQIAIMTFKDYFYCWNTFVENHETFFKAWEGHNSVRLSRLRRL 180
DB 167 KKLKAEAGCKLRIMKPDQFEYIWNQFVEQEGESKAFEPWEDIQENFLYYEEKLADIL 223

RESULT 11
Q9Y235
ID Q9Y235 PRELIMINARY; PRT; 224 AA.
AC Q9Y235
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE DJ34B21.2 (Putative novel protein similar to PART of APOBEC1
DE (PHORBOLIN 1, apolipoprotein B mRNA editing protein)).
GN DJ34B21.2 OR APOBEC-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=HEART;
RA Peck A.;
RT "APOBEC-2, a cardiac and skeletal muscle specific member of the
RT cytidine deaminase supergene family.";
RL submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
```

```

RN  SEQUENCE FROM N.A.
RP  TISSUE=HEART;
RX  MEDLINE=99333690; PubMed=10403781;
RA  Liao W., Hong S.-H., Chang B.-J., Rudolph F.B., Clark S.C., Chan L.;
RT  "APOBEC-2, a cardiac- and skeletal muscle-specific member of the
RL  cytidine deaminase supergene family.";
DR  Blochem. Biophys. Res. Commun. 260:398-404(1999).
DR  EMBL; AF031778; CAB44740.1; -.
KW  Lipoprotein.
SQ  SEQUENCE 224 AA; 25703 MW; CA0905AFAA8C8FA1 CRC64;

Query Match      25.2%; Score 274; DB 4; Length 224;
Best Local Similarity 36.2%; Pred. No. 8.6e-20;
Matches 64; Conservative 25; Mismatches 76; Indels 12; Gaps 5;

QY  11 FLYQFNVRWAKGRRTYCYVVKRDSATSFSLDFGLNKN-GCHVELLFLRYISDWD 69
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db  52 FKFQFRVEYSSGRNKFCLYVVEAOGKGVQASRGYLEDEHAAHAEEAFFNTILP-A 110
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY  70 LDPGRCYRYTWTFSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPE---GL 126
    |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db  111 FDPALRYNTWTVSSSPCAACADRIIKTSKTNLRLILLVGLFMWE-----EPEIQAL 166

QY  127 RLLHRAGVQIAIMTFKDYFCWNTFVENHE---RTFKAWBGLHNSVRLSRQLRRIL 180
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  167 KKLKEAGCKLRIMKPODFEVMQNFVEQEGESKAFQPWEDIQENFLYYEEKLADIL 223

RESULT 12
QY553
ID  QY553      PRELIMINARY;      PRT;      121 AA.
AC  QY553;
DT  01-NOV-1999 (TrEMBLrel. 12, Created)
DT  01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  BK150C2.10 (Putative novel Phorbolin 1 like protein) (Fragment).
GN  BK150C2.10.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Matthews L.;
RL  Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AL022318; CAB45273.1; -.
FT  NON_TER 1
FT  NON_TER 121
SQ  SEQUENCE 121 AA; 14309 MW; 97FC47DCD82B47 CRC64;

Query Match      24.9%; Score 270.5; DB 4; Length 121;
Best Local Similarity 45.8%; Pred. No. 9.7e-20;
Matches 55; Conservative 17; Mismatches 47; Indels 1; Gaps 1;

QY  61 FLRYISDWDLPGRCYRYVWTFSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRK 120
    ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  3 FLFWCDDILSPNTNIEVYTWYTSWSPCEGAEVAFELARHSNVNITFTARLYCWFWD 62

QY  121 ABPEGLRRLHRAVGQVIAIMTFKDYFCWNTFVENHRTFKAWBGLHNSVRLSRQLRRIL 180
    :||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  63 YQ-EGLCSLSQEGASVKINGKDFVSCWKNFVYSDDEPFKPKGLQTNFLLKRRLEIL 121

RESULT 13
QY550
ID  QY550      PRELIMINARY;      PRT;      103 AA.
AC  QY550;
DT  01-NOV-1999 (TrEMBLrel. 12, Created)
DT  01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT  01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE  BK150C2.6 (Putative novel protein similar to APOBEC1 (Apolipoprotein B

```

```

DE  mRNA editing protein) and Phorbolin) (Fragment).
GN  BK150C2.6.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Matthews L.;
RL  Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AL022318; CAB45276.1; -.
DR  InterPro; IPR002125; dCMP/cyt_deam.
DR  PROSITE; PS00903; CYT_DCMPE_DEAMINASES; UNKNOWN_1.
KW  Lipoprotein.
FT  NON_TER 1
FT  NON_TER 103
SQ  SEQUENCE 103 AA; 12146 MW; 5DC969AE3ED348D9 CRC64;

Query Match      23.7%; Score 257.5; DB 4; Length 103;
Best Local Similarity 50.0%; Pred. No. 1.7e-18;
Matches 49; Conservative 13; Mismatches 35; Indels 1; Gaps 1;

QY  55 CHVELLEFLRYISDWDLPGRCYRYVWTFSWSPCYDCARHVADFLRGNPNLSLRIFTARLY 114
    ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  7 CHAERCLSWFCDDILSPNTNIEVYTWYTSWSPCEGAEVAFELARHSNVNITFTARLY 66

QY  115 FCEDRKAEPEGLRRLHRAVGQVIAIMTFKDYFCWNTFV 152
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  67 YFWDTDYQ-EGLSLSQEGASVEIMGYKDFKFCWNTFV 103

RESULT 14
QY554
ID  QY554      PRELIMINARY;      PRT;      111 AA.
AC  QY554;
DT  01-MAY-2000 (TrEMBLrel. 13, Created)
DT  01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT  01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE  BK150C2.1 (Putative novel Phorbolin 1 like protein) (Fragment).
GN  BK150C2.1.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RA  Matthews L.;
RL  Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AL022318; CAB45269.1; -.
FT  NON_TER 1
FT  NON_TER 111
SQ  SEQUENCE 111 AA; 13006 MW; EF8922AEFFA7028 CRC64;

Query Match      23.1%; Score 251; DB 4; Length 111;
Best Local Similarity 46.4%; Pred. No. 8.2e-18;
Matches 51; Conservative 20; Mismatches 29; Indels 10; Gaps 3;

QY  76 YRYTWTFSWSPCYD--CARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPE---EGLRRLH 130
    ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  2 YRYTWTFSWSPCWGCGAEVAFELQENTHRLRIFAARIY-----DYDPLYKEALQMLR 56

QY  131 RAGVQIAIMTFKDYFCWNTFVENHRTFKAWBGLHNSVRLSRQLRRIL 180
    ||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db  57 DAGAQSVMYDEKHKWDTFVDHQGCPFPQPDGDLDEHSQALSGLRLAIL 106

RESULT 15
QY554
ID  QY554      PRELIMINARY;      PRT;      121 AA.
AC  QY554;
DT  01-NOV-1999 (TrEMBLrel. 12, Created)
DT  01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  BK150C2.9 (Putative novel Phorbolin 1 like protein) (Fragment).

```

```

GN BK150C2.9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022318; CAB45272.1; -.
FT NON_TER 1
FT NON_TER 121
SQ SEQUENCE 121 AA; 14714 MW; B8FE133AACBE6D59 CRC64;

Query Match      22.1%; Score 240.5; DB 4; Length 121;
Best Local Similarity 39.3%; Pred. No. 1e-16;
Matches 48; Conservative 19; Mismatches 54; Indels 1; Gaps 1;

QY 59 LFLFLYISDWLDLPGRCYRVTFWTSWSPCYDCARHVADFLRGPNLRLIFRTARYFCED 118
Db 1 MCFLSWFCGNRLPANRRFOITWFSWNPCLEPCVVVKTKFLAEHPNVTLTISAARLYYRD 60

QY 119 RKAPEGLRLRHAGVQIAIMTFKDYFYCWNTFVENHERTEKAWGLHENSRLSRQLRR 178
Db 61 RDWR-WVLLRLHKGARVKINDYEDFAFCWENFVCNCGQPPMPWKFDNDNYASLHRTIKE 119

QY 179 IL 180
Db 11
Db 120 IL 121

```

Search completed: June 14, 2003, 18:11:07  
Job time : 80 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2003, 18:08:03 ; Search time 26 Seconds  
(without alignments)  
224.067 Million cell updates/sec

Title: US-09-966-880A-8  
Perfect score: 1086  
Sequence: 1 MDSLLMNRKFLYQFKVNRW.....ILLPLYEVDLDAFRTLGL 198

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	369.5	34.0	190	1 US-08-816-241-1	Sequence 1, Appli
2	369.5	34.0	190	3 US-09-128-395-1	Sequence 1, Appli
3	271	25.0	222	1 US-08-687-895-1	Sequence 1, Appli
4	271	25.0	222	2 US-09-040-482-1	Sequence 1, Appli
5	245	22.6	116	1 US-08-687-895-3	Sequence 3, Appli
6	245	22.6	116	1 US-08-816-241-3	Sequence 3, Appli
7	245	22.6	116	2 US-09-040-482-3	Sequence 3, Appli
8	245	22.6	116	3 US-09-128-395-3	Sequence 3, Appli
9	216.5	19.9	236	1 US-08-687-895-4	Sequence 4, Appli
10	216.5	19.9	236	1 US-08-816-241-4	Sequence 4, Appli
11	216.5	19.9	236	2 US-09-040-482-4	Sequence 4, Appli
12	216.5	19.9	236	3 US-09-128-395-4	Sequence 4, Appli
13	210	19.3	236	1 US-08-158-682A-4	Sequence 4, Appli
14	198	18.2	229	1 US-08-158-682A-2	Sequence 2, Appli
15	198	18.2	229	1 US-08-015-203-2	Sequence 2, Appli
16	198	18.2	229	1 US-08-687-895-5	Sequence 5, Appli
17	198	18.2	229	1 US-08-816-241-5	Sequence 5, Appli
18	198	18.2	229	2 US-09-040-482-5	Sequence 5, Appli
19	198	18.2	229	3 US-09-128-395-5	Sequence 5, Appli
20	81	7.5	440	4 US-09-092-315-3	Sequence 3, Appli
21	80	7.4	476	4 US-09-092-315-5	Sequence 5, Appli
22	77.5	7.1	402	6 5352575-5	Patent No. 5352575
23	75.5	7.0	343	1 US-08-180-209B-56	Sequence 56, Appli
24	75.5	7.0	343	4 US-08-474-853-56	Sequence 56, Appli
25	75.5	7.0	343	4 US-09-166-205B-56	Sequence 56, Appli
26	75.5	7.0	343	5 PCT-US94-02629-56	Sequence 56, Appli
27	73.5	6.8	533	1 US-08-488-305A-6	Sequence 6, Appli

28	72.5	6.7	366	6	5470718-4	Patent No. 5470718
29	72	6.6	1094	2	US-08-680-326-40	Sequence 40, Appli
30	72	6.6	1400	1	US-08-080-255-7	Sequence 7, Appli
31	72	6.6	1400	3	US-08-465-713-7	Sequence 7, Appli
32	72	6.6	1400	5	PCT-US93-05857-7	Sequence 7, Appli
33	72	6.6	3969	4	US-08-061-376-5	Sequence 5, Appli
34	70.5	6.5	528	6	5475095-2	Patent No. 5475095
35	69.5	6.4	533	4	US-09-385-259-2	Sequence 2, Appli
36	69.5	6.4	533	4	US-09-645-370-2	Sequence 2, Appli
37	69.5	6.4	1727	2	US-08-477-451-10	Sequence 10, Appli
38	69	6.4	486	4	US-09-092-315-2	Sequence 2, Appli
39	68	6.3	425	4	US-09-092-315-6	Sequence 6, Appli
40	68	6.3	1176	1	US-07-828-788A-12	Sequence 12, Appli
41	68	6.3	1176	1	US-08-356-034-2	Sequence 2, Appli
42	68	6.3	1176	3	US-08-933-891-2	Sequence 2, Appli
43	68	6.3	1176	5	PCT-US92-11337-12	Sequence 12, Appli
44	68	6.3	1179	1	US-08-040-751-1	Sequence 1, Appli
45	67	6.2	207	4	US-09-415-522-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-816-241-1  
; Sequence 1, Application US/08816241  
; Patent No. 5804185  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/816,241  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0239 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 190 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: PROSTUT09  
; CLONE: 1646823  
; US-08-816-241-1

Query Match 34.0%; Score 369.5; DB 1; Length 190;  
Best Local Similarity 44.9%; Pred. No. 1.2e-35;  
Matches 79; Conservative 24; Mismatches 64; Indels 9; Gaps 4;





```

; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/040,482
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/687,895
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 222 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: MUSCNOT1
; CLONE: 57953
;
US-09-040-482-1

Query Match 25.0%; Score 271; DB 2; Length 222;
Best Local Similarity 35.8%; Pred. No. 6e-24;
Matches 63; Conservative 25; Mismatches 76; Indels 12; Gaps 5;

QY 11 FLYQFNWAKGRRETYLYVYVVRDSDTSFSLDFGYLRNKN-GCHVELLFLRYISDWD 69
| : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : |
52 FKFPNRYVSSGRNKTFLCYVVEAQGGQVQASRGYLEDEHAAHAEAFNTILP-A 110
| : : : : : | : : : : : | : : : : : | : : : : : |
70 LPDGRYRVTWTSWSPCYDCARHVADFLRGNPNLSRIPTARLYFCEDRKAEP---GL 126
| : : : : : | : : : : : | : : : : : | : : : : : |
111 FDPALRYNTWTVSSSPCAACADRIKXKTLKLNRLLLVGLRFWME----EPEIQAA 166
| : : : : : | : : : : : | : : : : : | : : : : : |
127 RRLHRAGVOIAITMFKDYFCVNTFVENHE---RTFKAMEGLHNSVRLSRQLRI 179
| : : : : : | : : : : : | : : : : : | : : : : : |
167 KKLKAGCKLRINKPDQFEYVQNFVEQEGESKAFQPDIEDIYFYEKLADI 222
| : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 5
US-08-687-895-3
; Sequence 3, Application US/08687895
; Patent No. 5747319
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,895
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 116 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 436941
;
US-08-687-895-3

Query Match 22.6%; Score 245; DB 1; Length 116;
Best Local Similarity 46.4%; Pred. No. 2.8e-21;
Matches 51; Conservative 17; Mismatches 32; Indels 10; Gaps 3;

QY 76 YRVTFWTSWSPCYD--CARHVADFLRGNPNLSRIPTARLYFCEDRKAEP---EGLRRLH 130
| : : : : : | : : : : : | : : : : : | : : : : : |
| : : : : : | : : : : : | : : : : : | : : : : : |
7 YRVTFWTSWSPCFSGWCAGEVRAFLQENTHVRUPIFAARIY-----DYDPLYKEALQMLR 61
| : : : : : | : : : : : | : : : : : | : : : : : |

QY 131 RAGVQIAITMFKDYFCVNTFVENHRTFKAMEGLHNSVRLSRQLRI 180
| : : : : : | : : : : : | : : : : : | : : : : : |
62 DAGAQVSIPTYDEFYCWDTFVYRQGCFFQPDWGLEHSHQALSGLRLRAIL 111
| : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 6
US-08-816-241-3
; Sequence 3, Application US/08816241
; Patent No. 5804185
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,241
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0239 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555

```

TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 116 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 436941  
US-08-816-241-3

Query Match 22.6%; Score 245; DB 1; Length 116;  
Best Local Similarity 46.4%; Pred. No. 2.8e-21;  
Matches 51; Conservative 17; Mismatches 32; Indels 10; Gaps 3;  
QY 76 YRVTFWTSWSPCYD--CARHVADFRLGNPNLSLRIETARLYFCEDRKAEP---EGLRRLH 130  
DB 7 YRVTFWTSWSPCYD--CARHVADFRLGNPNLSLRIETARLYFCEDRKAEP---EGLRRLH 130  
QY 131 RAGVQIAIMTFKDYFCWNTFVENHIERTFKAWEGLEHNSVRLSRQLRRIL 180  
DB 62 DAGAQVSIIMTYDEFEYCWDTFVYRQCPCFPQWDGLEHNSQALSGRLRAIL 111

RESULT 7  
US-09-040-482-3  
Sequence 3, Application US/09040482  
Patent No. 5916556  
GENERAL INFORMATION:  
APPLICANT: Au-Young, Janice  
APPLICANT: Hawkins, Phillip R.  
APPLICANT: Hillman, Jennifer L.  
TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: U.S.  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/040,482  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/687,895  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0109 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 116 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 436941  
US-09-040-482-3

Query Match 22.6%; Score 245; DB 2; Length 116;

Best Local Similarity 46.4%; Pred. No. 2.8e-21;  
Matches 51; Conservative 17; Mismatches 32; Indels 10; Gaps 3;  
QY 76 YRVTFWTSWSPCYD--CARHVADFRLGNPNLSLRIETARLYFCEDRKAEP---EGLRRLH 130  
DB 7 YRVTFWTSWSPCYD--CARHVADFRLGNPNLSLRIETARLYFCEDRKAEP---EGLRRLH 130  
QY 131 RAGVQIAIMTFKDYFCWNTFVENHIERTFKAWEGLEHNSVRLSRQLRRIL 180  
DB 62 DAGAQVSIIMTYDEFEYCWDTFVYRQCPCFPQWDGLEHNSQALSGRLRAIL 111

RESULT 8  
US-09-128-395-3  
Sequence 3, Application US/09128395  
Patent No. 6087108  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Goli, Surya K.  
TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/128,395  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/816,241  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0239 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 116 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: GenBank  
CLONE: 436941  
US-09-128-395-3

Query Match 22.6%; Score 245; DB 3; Length 116;  
Best Local Similarity 46.4%; Pred. No. 2.8e-21;  
Matches 51; Conservative 17; Mismatches 32; Indels 10; Gaps 3;  
QY 76 YRVTFWTSWSPCYD--CARHVADFRLGNPNLSLRIETARLYFCEDRKAEP---EGLRRLH 130  
DB 7 YRVTFWTSWSPCYD--CARHVADFRLGNPNLSLRIETARLYFCEDRKAEP---EGLRRLH 130  
QY 131 RAGVQIAIMTFKDYFCWNTFVENHIERTFKAWEGLEHNSVRLSRQLRRIL 180  
DB 62 DAGAQVSIIMTYDEFEYCWDTFVYRQCPCFPQWDGLEHNSQALSGRLRAIL 111

RESULT 9  
US-08-687-895-4

```

; Sequence 4, Application US/08687895
; Patent No. 5747319
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,895
; FILING DATE: Filed Herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0109 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 236 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1177798
; US-08-687-895-4

Query Match 19.9%; Score 216.5; DB 1; Length 236;
Best Local Similarity 34.9%; Pred. No. 1.7e-17;
Matches 44; Conservative 31; Mismatches 42; Indels 9; Gaps 4

QY 35 RROSATSFSLDFGYLR-----NKNGC-HVLELLF-RYISDWLLDPGRCYRVTWFTWSP 86
Db 33 RKACALLEYIKWGMRSIKWRSSGNTNTHVEVNFKKFTSERDFHPSICSTITWFLWSP 92
QY 87 CYDCARHVADFLRGNPNLSLRIFATRYLFCEDRAEPEGLRRLRHAGVQIAIMTFKDYFY 146
Db 93 CWESQAIREFLSRHPGVTLYIYARLFWHMDQ-NRQGLRLVNSGVTIQIMRASEYH 151
QY 147 CWNTFV 152
Db 152 CWRNFV 157

RESULT 10
US-08-816-241-4
; Sequence 4, Application US/08816241
; Patent No. 5804185
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goll, Surya K.
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA

```







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 14, 2003, 18:09:18 ; Search time 21 seconds  
(without alignments)  
1008.062 Million cell updates/sec

Title: US-09-966-880A-8  
Perfect score: 1086  
Sequence: 1 MDSLNNRRKFLYQFNVRW.....ILLPLYEVDLDRFAFTLGL 198

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 408643 seqs, 106915682 residues

Total number of hits satisfying chosen parameters: 408643

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1086	100.0	198	9	US-09-966-880A-8
2	1008	92.8	198	9	US-09-966-880A-8
3	390	35.9	384	10	US-09-729-674-174
4	369.5	34.0	222	10	US-09-925-300-1639
5	211	19.4	229	9	US-09-966-880A-36
6	84	7.7	51	10	US-09-864-761-38853
7	81	7.5	440	9	US-10-120-319-3
8	80	7.4	476	9	US-10-120-319-5
9	80	7.4	476	10	US-09-733-524-15
10	77.5	7.1	663	9	US-10-080-960-14
11	77	7.1	790	9	US-10-153-668-164
12	75.5	7.0	382	9	US-09-847-208-25
13	75	6.9	330	9	US-10-265-593-4
14	71.5	6.6	529	9	US-10-053-192-5
15	71	6.5	440	9	US-09-910-186A-8
16	69	6.4	485	10	US-09-733-524-2
17	69	6.4	486	9	US-10-120-319-2
18	69	6.4	2015	9	US-10-066-551-1
19	68.5	6.3	295	9	US-10-042-894A-14

20	68.5	6.3	341	9	US-09-764-868-1041	Sequence 1041, Ap
21	68.5	6.3	491	10	US-09-925-302-762	Sequence 762, App
22	68	6.3	424	10	US-09-733-524-16	Sequence 16, Appl
23	68	6.3	425	9	US-10-120-319-6	Sequence 6, Appli
24	67	6.2	454	9	US-10-120-319-8	Sequence 8, Appli
25	67	6.2	454	10	US-09-733-524-18	Sequence 18, Appl
26	67	6.2	479	10	US-09-733-524-17	Sequence 17, Appl
27	67	6.2	602	10	US-09-841-132-495	Sequence 495, App
28	66.5	6.1	287	10	US-09-864-761-42836	Sequence 42836, A
29	66.5	6.1	386	10	US-09-839-497A-5	Sequence 5, Appli
30	66	6.1	440	10	US-09-733-524-3	Sequence 3, Appli
31	65.5	6.0	239	9	US-10-219-220-287	Sequence 287, App
32	65.5	6.0	1045	10	US-09-815-242-10617	Sequence 10617, A
33	65.5	6.0	4636	10	US-09-835-996A-33	Sequence 33, Appl
34	65	6.0	410	9	US-09-987-021-2	Sequence 2, Appli
35	65	6.0	410	10	US-09-957-485-2	Sequence 2, Appli
36	65	6.0	602	10	US-09-841-132-565	Sequence 565, App
37	65	6.0	637	9	US-10-041-007-24	Sequence 24, Appl
38	65	6.0	637	9	US-09-900-797-58	Sequence 58, Appl
39	65	6.0	637	10	US-09-887-586A-58	Sequence 58, Appl
40	65	6.0	637	10	US-09-903-012-58	Sequence 58, Appl
41	65	6.0	905	9	US-10-114-893-127	Sequence 127, App
42	65	6.0	1005	10	US-09-802-127-5	Sequence 5, Appli
43	64.5	5.9	250	9	US-09-880-748-1480	Sequence 1480, Ap
44	64.5	5.9	373	9	US-10-125-540-353	Sequence 353, App
45	64.5	5.9	373	9	US-10-103-313-437	Sequence 437, App

ALIGNMENTS

RESULT 1  
US-09-966-880A-8  
; Sequence 8, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966, 880A  
; PRIOR FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 198  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-966-880A-8  
  
Query Match 100.0%; Score 1086; DB 9; Length 198;  
Best Local Similarity 100.0%; Pred. No. 3e-108;  
Matches 198; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
QY 1 MDSLNNRRKFLYQFNVRWAKGRRETYLCYVVKRDSATSFSLDFGLNKNKGVHLL 60  
Db 1 MDSLNNRRKFLYQFNVRWAKGRRETYLCYVVKRDSATSFSLDFGLNKNKGVHLL 60  
QY 61 FLRYISDWLDLPGRCYRVTFWFTSWSPCYDCARHVAFLRGNPNLSLRIFATRLYFCSDRK 120  
Db 61 FLRYISDWLDLPGRCYRVTFWFTSWSPCYDCARHVAFLRGNPNLSLRIFATRLYFCSDRK 120  
QY 121 ASPEGLRRLRHAGVQIAIMTFKDYFYCWNTFVENHRTFKAWEGHNSVRLSRQLRRL 180  
Db 121 ASPEGLRRLRHAGVQIAIMTFKDYFYCWNTFVENHRTFKAWEGHNSVRLSRQLRRL 180

QY 181 LPLYEVDDL RDAFRTLGL 198  
 |||||  
 Db 181 LPLYEVDDL RDAFRTLGL 198

## RESULT 2

US-09-966-880A-2  
; Sequence 2, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:

; APPLICANT: Honjo, Tasuku  
 ; APPLICANT: Muramatsu, Masamichi  
 ; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE

```

; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A

```

; CURRENT FILING DATE: 2001-09-28  
 ; PRIOR APPLICATION NUMBER: PCT/JP00/01918

; PRIOR FILING DATE: 2000-03-28  
 ; PRIOR APPLICATION NUMBER: JP 11-371382

; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999

; PRIORITY FILING DATE: 1999-06-24  
 ; PRIORITY APPLICATION NUMBER: JP 11-87192

; PRIORITY FILING DATE: 1999-03-29  
 ; NUMBER OF SEQ ID NOS: 36

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2

```

```

; SEQ ID NO: 2
; LENGTH: 198
; TYPE: DDT

```

TYPE: PRI  
; ORGANISM:  
US-09-966-880

03-09-300-880A-2  
Quorum Match

Query Match	92.8%;	Score 1008;	DB 9;	Length 198;
Best Local Similarity	92.9%;	Pred. No. 6.5e-100;		
Matches 183;	Conservative	6;	Mismatches	8;
			Indels	

QY  
1 MDSLLMNRKFLYQFKNVRWAKGRRETYLCYVVRDRSATSFSLDFGYLRNKGCHVELL 60

Db 1 MD5LLMKQKKFELYHFKNVRWAKGRHETLYCYVVKRRDSATSCSLDFGHLRNKSGCHVELL 60

QY 61 FLRYISDWDLDPGRCYRVTWETSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRK 120

Db 61 FLRYISDWLDLPGRCYRVTWFTSWSPCYDCARHVAEFLRWNPENLSLRIFARTARLYFCEDRK 120

QY 121 AEPEGLRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWGLHENSRLSRQLRRL 180

D6 121 AEPEGLRLHRAGVQIGIMFKDYFCYCNTEFVENPERTFKAMEGLHENSVRTQLRRIL 180

Qy 181 LPLYEVDDLDRDAFTLG 197

Db 181 LPLYEVDLDRDAFRMLG 197

RESULT 3  
US-09-729-674-174

; Sequence 174, Application US/09729674  
; Patent No. US20010039335A1

; GENERAL INFORMATION:  
 ; APPLICANT: Jacobs, Kenneth

APPLICANT: McCoy, John M.  
APPLICANT: Lavallie, Edward R.

APPLICANT: Collins-Racie, Lisa A.  
APPLICANT: Evans Cheryl

APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
APPLICANT: Moseley, Maurice

; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; ADDITIONAL: Steininger, Robert

APPLICANT: Steininger II, Robert J.  
APPLICANT: Spaulding, Vikki

APPLICANT: Wong, Gordon G.  
APPLICANT: Clark, Hilary

; APPLICANT: Fecthel, Kim  
; APPLICANT: Genetics Institute, Inc.

```

: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
:
: FILE REFERENCE: 6055-64X
:
: CURRENT APPLICATION NUMBER: US/09/729,674
:
: CURRENT FILING DATE: 2000-12-04
:
: PRIOR APPLICATION NUMBER: 09/539,330
:
: PRIOR FILING DATE: 2000-03-30
:
: NUMBER OF SEQ ID NOS: 283
:
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 174
:
: LENGTH: 384
:
: TYPE: PRT
:
: ORGANISM: Homo sapiens
:
: US-09-729-674-174

```

Query Match	35.9%	Score 390;	DB 10;	Length 384;
Best Local Similarity	44.9%;	Pred. No. 1.6e-33;		
Matches 83; Conservative	31;	Mismatches 59;	Indels 12;	Gaps 3;
QY	6	MNRKRFLYQKFNVRWAKGRRETYLCYVVVKRRDSATSFSLD--FCYLRNK-----NGC	55	
Db	197	MDPTTFTEFNNEPWPVGRHETIYCYEVERMHNDTWVLLNORRGFLCNOAPKHGFLGGR	256	
QY	56	HVLELLFURYTSDMDLDGRCYRYVTWFTSWSPCYDCARHVADFLRGNPNLSURITARIYLF	115	
Db	257	HAELCFELDVFPFKWLDLDQDYRYVTCTFSWSPCFSCAQEMAKFISKNKHVSLICITARIY-	315	
QY	116	CEDRKABEPEGLRRLRHAGVOLAINTPKDYFCWNTEFVENHERTFKAMEGLHENSVRLSRQ	175	
Db	316	-DQGRCEQEGRLTAEAGAKISIMTYSEFFKHCWDTFVDHGQCFQPQPMWGLDGLDHSQDL	374	
QY	176	LRRIL 180		
Db	375	LRAIL 379		

## RESULT 4

```

US-09-925-300-1639
; Sequence 1639, Application US/09925300
; Patent NO. US2020151691A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Protei
; FILE REFERENCE: Pa101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1639
; LENGTH: 222
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-925-300-1639

```

	Query Match Best Local Matches	34.0% Similarity 79: Conservative	Score 369.5; Pred. No. 1.3e-31; Mismatches 24;	DB 10; Length 222; Indels 9;
QY	11	FLYQFNVRWAKGRRETYLCYVVK--RRDSATSFSLDFGYLRN----	KNCSHVLELLFLRY	64
Db	49	FYQFNLEWANDRNENETWLCFTVEGKKRSVSVWKT--GVFRNQVDSETHCAERCSFLW		106
QY	65	ISDWDLDPGRCYRVWTFNTSPCYDCARHVADELRGNPNLSLRIFITARLYFYCEDRKAEP		124
Db	107	ECDDILSPNTKQVWYTWTSVPCPCDCAGEVAEFLARHSNVNLIIFTARLYFYQ-YPYQYE		165
QY	125	GLURLHRAGVQIAIMTFKDYFYCWNTFEVNHERTFYKAWEGCLHNSVRLSRQLRRL		180
Db	166	GLRSLSQEGVAVIEMDYEDQFYCWENFVYNDNPPFKPKGLKTNFRLKKRLRESL		221



```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 38853
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL022318.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.1
; OTHER INFORMATION: EST_HUMAN HIT: BF237537.1, EVALUATE 2.00e-15
US-09-864-761-38853

Query Match 7.7%; Score 84; DB 10; Length 51;
Best Local Similarity 40.0%; Pred. No. 0.068;
Matches 20; Conservative 8; Mismatches 20; Indels 2; Gaps 2;

QY 5 LMNRKFLYQFKNVWAKGRRETYLCYVVRDSDATSFSLD--FGYLANK 52
Db 2 LMDPDTFTFNFDPLVLRRTYLCYVEVERLDGTWVLMDOHMGFLCNE 51

RESULT 7
US-120-319-3
; Sequence 3, Application US/10120319
; Patent No. US20020164749A1
; GENERAL INFORMATION:
; APPLICANT: Taylor, Diane E.
; APPLICANT: Ge Zhongming
; TITLE OF INVENTION: ALPHA-1, 3-FUCOSYLTRANSFERASE
; FILE REFERENCE: 07254/049001
; CURRENT APPLICATION NUMBER: US/10/120,319
; CURRENT FILING DATE: 2002-04-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/092,315
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/048,857
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-06-06
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-120-319-3

```



Query Match 7.1%; Score 77.5; DB 9; Length 663;  
Best Local Similarity 22.5%; Pred. No. 7.8;  
Matches 42; Conservative 36; Mismatches 94; Indels 15; Gaps 6;

QY 16 KNRWAKGRRETYLCVVVKRRDSATSFSLDFGLRNKNGCHVLELLFLRYSID---WDLP 72  
DB 249 EKIQKSGDDATLPSRFLPKDGTGTRIGDLAPQDMKKVCHLALIELTALYDVLGIELKQ 308  
QY 73 GRCYRVTWFTSPCYDCARHVADFLRGNPNLSL-RIFTARLYFCEDRKAEPGLRLRHR 131  
DB 309 ORAVKIKTRDGLFVPLTALLDQDQKVPGMRIPLIFOKLISRIBERGLEYBGLRIPG 368  
QY 132 AGVQIAIMTFKDYFCWNTFVENHRTFKAWGL--HENSVRSLRQLRLLPLYEVDL 189  
DB 369 AAIRIKNL-----COELEAKFYEGTFN-WESVKQHDAAALLKLFIRELPQPLLSVEYL 420  
QY 190 RDAFRTL 196  
DB 421 K-AFAQV 426

RESULT 11  
US-10-153-668-164  
; Sequence 164; Application US/10153668  
; Publication No. US20030092616A1  
; GENERAL INFORMATION:  
; APPLICANT: HONDA, Goichi  
; APPLICANT: MATSUDA, Akio  
; APPLICANT: MURAMATSU, Shuji  
; APPLICANT: ISHIZAWA, Kenya  
; TITLE OF INVENTION: STAT6 Activating Gene  
; FILE REFERENCE: 1254-0207P  
; CURRENT APPLICATION NUMBER: US/10/153,668  
; CURRENT FILING DATE: 2002-05-24  
; PRIOR APPLICATION NUMBER: US 60/293,172  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: US 60/316,031  
; PRIOR FILING DATE: 2001-08-31  
; PRIOR APPLICATION NUMBER: US 60/328,403  
; PRIOR FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: JP 2001-157043  
; PRIOR FILING DATE: 2001-05-25  
; PRIOR APPLICATION NUMBER: JP 2001-260681  
; PRIOR FILING DATE: 2001-08-30  
; PRIOR APPLICATION NUMBER: JP 2001-313175  
; PRIOR FILING DATE: 2001-10-10  
; NUMBER OF SEQ ID NOS: 488  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 164  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-153-668-164

Query Match 7.1%; Score 77; DB 9; Length 790;  
Best Local Similarity 23.2%; Pred. No. 11;  
Matches 51; Conservative 23; Mismatches 66; Indels 80; Gaps 11;

QY 29 LCYVVKRRDSATSFSL-----DFGY-----LRNKNCHVLELLFLRYSIDWLDLPGRC 75  
DB 584 LCSVLKASGLQSLSDATSFASPDQGLVLOTLKEYN-----LALKRLSFHDMNLAD 637  
QY 76 YRVTFP-----TSWSPCYDCARHVADEL-----RGNPNL-SLRIFTARL----- 113  
DB 638 QSEVLELLONLTLQELTFSCFLFEKRPQOFLPEWVAAMKGNSTLGLRLPGNRGNAGL 697  
QY 114 -----YFCEDRKA-----EPEGL-----RRLHRAGVQIALMTFKDYFCWNTFV 152  
DB 698 LALADVSESDSSSLQQLDISSNCIKPDGLLEFAKLER-----WGRGA 741  
QY 153 ENHRTFKAWGLHENSVRSLRQLRLLPLYEVDL 192  
DB 742 FGHLRLFQNW--LDQDAVTAREAIRRLRATCHVWSDSDS 779

RESULT 12  
US-09-847-208-25  
; Sequence 25; Application US/09847208  
; Publication No. US20030082190A1  
; GENERAL INFORMATION:  
; APPLICANT: Saxon, Andrew  
; APPLICANT: Zhang, Ke  
; APPLICANT: Zhu, Daocheng  
; TITLE OF INVENTION: FUSION MOLECULES AND TREATMENT OF  
; TITLE OF INVENTION: I98-MEDIATED ALLERGIC DISEASES  
; FILE REFERENCE: UC67.002A  
; CURRENT APPLICATION NUMBER: US/09/847,208  
; CURRENT FILING DATE: 2001-05-01  
; NUMBER OF SEQ ID NOS: 177  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 25  
; LENGTH: 382  
; TYPE: PRT  
; ORGANISM: Apis mellifera (Honeybee)  
US-09-847-208-25

Query Match 7.0%; Score 75.5; DB 9; Length 382;  
Best Local Similarity 21.6%; Pred. No. 6.5;  
Matches 36; Conservative 21; Mismatches 51; Indels 59; Gaps 7;

QY 81 FTSWSPCYDCARHVADFLRGNPNLSLRIFTARLYFCEDRKAEPGLRLRHRAGVQIAIMT 140  
DB 144 FESWRPIF---RONWASLQPYKKLSVEVVRREHPFWDQVEQEAQRFRFEKYGQLFMEET 200  
QY 141 FK-----DYFYCWN-----TFVENHRT-----FK 160  
DB 201 LKAAKMRPAANWGYAYPYCYNLTPNPSAQACEATTMQENDKMSWLFESDVLPSVYL 260  
QY 161 AWE-----GLHENSVRLSROL---RRILLPL--YEVDLDR 191  
DB 261 RWNLTSGERVLGVRVKELARITARQMTTSRKKVLPLYWYKYQDQRRD 307

RESULT 13  
US-10-265-593-4  
; Sequence 4; Application US/10265593  
; Publication No. US20030049789A1  
; GENERAL INFORMATION:  
; APPLICANT: LADUNGA, Steven et al  
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS,  
; TITLE OF INVENTION: USES THEREOF  
; FILE REFERENCE: CL000818CON  
; CURRENT APPLICATION NUMBER: US/10/265,593  
; CURRENT FILING DATE: 2002-10-08  
; PRIOR APPLICATION NUMBER: 09/708,701  
; PRIOR FILING DATE: 2000-09-11  
; PRIOR APPLICATION NUMBER: 60/243,467  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 4  
; LENGTH: 330  
; TYPE: PRT  
; ORGANISM: Chick  
US-10-265-593-4

Query Match 6.9%; Score 75; DB 9; Length 330;  
Best Local Similarity 23.6%; Pred. No. 6.1;  
Matches 42; Conservative 18; Mismatches 74; Indels 44; Gaps 8;

QY 1 MDSLNRNRKFLYQFKNVRWAKORRETYLCYVVKRRDSATSFSLDFGLRNKNGCHVLELL 60  
DB 52 MEAISMSALECOYQFRFRMNCTLEGYRASLLKRGFKETAFL-----YATSSAG----- 101  
QY 61 FLRYISDWLDLPGRCYRVT-----WFTSWSPCYDCARH-----VADFLRGNPNL 104

Db . 102 -LTHAMKACAGNRCTCDEADPLENREAW--QWGGCGDNLYSNKFKEFLGRKPNK 158  
Qy 105 SLRIETARLYFCEDR-----KAEPEGLRLRLHRAGVQIAIMTFKDYFCVNTFVENHE 156  
Db 159 DLR---ARVDHNNLVGMKVIKAGVETTKCHGVSSGCTVRT-----CNRQLSPFHE 207

## RESULT 14

US-10-053-192-5

; Sequence 5, Application US/10053192  
; Publication No. US20030087336A1  
; GENERAL INFORMATION:  
; APPLICANT: BACHMANN, Heinrich  
; APPLICANT: BRUGGER, Roland  
; APPLICANT: FRIEDLEIN, Arno M  
; APPLICANT: WIRTZ, Gabriele M  
; APPLICANT: WOGGON, Wolf-Dietrich  
; APPLICANT: WYSS, Adrian  
; APPLICANT: WYSS, Markus  
; TITLE OF INVENTION: BETA, BETA-CAROTENE 15,15'-DIOXYGENASES, NUCLEIC ACID  
; FILE REFERENCE: B.B-CAROTENE 15,15'-DIOXYGENASES, ...  
; CURRENT APPLICATION NUMBER: US/10/053,192  
; CURRENT FILING DATE: 2002-01-15  
; PRIOR APPLICATION NUMBER: 103382.0  
; PRIOR FILING DATE: 1999-02-22  
; NUMBER OF SEQ ID NOS: 10  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 529  
; TYPE: PRT  
; ORGANISM: BOVINE  
US-10-053-192-5

Query Match 6.6%; Score 71.5; DB 9; Length 529;  
Best Local Similarity 20.8%; Pred. No. 26;  
Matches 31; Conservative 25; Mismatches 42; Indels 51; Gaps 9;  
Qy 57 VELLFLRYISDWLDPCRCYRVVTFWTSWSPCYDCAR-----HVADEFLRGNPNLSLRI 108  
Db 246 VKINLFKFLSSWSL-----WGANYMDCPFSTNETMGVWLHIADKKR----- 287  
Qy 109 FTARLYFCEDRKAEPEGLRLRLHRAGVQIAIMTFKDYF-----CWN--TFVENHERTFK 160  
Db 288 ---KKYLNKKYTSFNL--FHH-----INTYEDNGFLIVDLCCWKGEFVYNYFTLYL 336  
Qy 161 A-----WEGLHENSRLSR-OLRRILLPL 183  
Db 337 ANLRNWEVVKKARKAPQEVRYVLP 365

## RESULT 15

US-09-910-186A-8

; Sequence 8, Application US/0910186A  
; Publication No. US20030009025A1  
; GENERAL INFORMATION:  
; APPLICANT: U.S. Army Medical Research & Material Command  
; TITLE OF INVENTION: RECOMBINANT VACCINE AGAINST BOTULINUM  
; FILE REFERENCE: A33626-A 067252.0107  
; CURRENT APPLICATION NUMBER: US/09/910,186A  
; CURRENT FILING DATE: 2001-07-20  
; PRIOR APPLICATION NUMBER: PCT/US00/12890  
; PRIOR FILING DATE: 2000-05-12  
; PRIOR APPLICATION NUMBER: 09/611,419  
; PRIOR FILING DATE: 2000-07-06  
; PRIOR APPLICATION NUMBER: 60/133,865  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,866  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,867  
; PRIOR FILING DATE: 1999-05-12

; PRIOR APPLICATION NUMBER: 60/133,868  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,869  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 60/133,873  
; PRIOR FILING DATE: 1999-05-12  
; PRIOR APPLICATION NUMBER: 08/123,975  
; PRIOR FILING DATE: 1993-09-21  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 440  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic construct based on Clostridium botulinum  
; OTHER INFORMATION: sequence  
US-09-910-186A-8

Query Match 6.5%; Score 71; DB 9; Length 440;  
Best Local Similarity 20.8%; Pred. No. 23;  
Matches 35; Conservative 26; Mismatches 49; Indels 58; Gaps 9;  
Qy 5 LMNRRKELYQFKNVWAKGRRETYL--CYVVKRRDSATSPS-----LDFGYLRN 51  
Db 270 ILTRSKYNQNSKYINY----RDLYIGEKFIRRKNSQSINDDIVRKEDYIYLDFFNLQ 325  
Qy 52 KNGCHV-----ELLFLRYISDWLDPCRCYRVVTFWTSW--SPCYDCARHVADFLRGN 101  
Db 326 ENRVYKYKPKKEEKLFLAPISDS-----ELNTIOIKEYDEQPTVSC----- 370  
Qy 102 PNLSLRIETARLYFCEDRKAEPE-----GLRRLHRAGVQIAIMTFKDYF 145  
Db 371 -----QLLFKKDEESTDEIGLIGIHRFYESG--IVFEYKDYF 406

Search completed: June 14, 2003, 18:12:57  
Job time : 23 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: June 19, 2003, 11:55:51 ; Search time 2485 Seconds

(without alignments)  
2318.855 Million cell updates/sec

Title: us-09-966-880a-8

Perfect score: 1086

Sequence: 1 MDSLMMNRKFLYQFKVRW.....ILLPLYEVDDLDRDAFRTLGL 198

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO\_spool/US09966880/runat\_14062003\_175645\_16271/app\_query.fasta\_1.391  
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09966880.@CGN\_1\_1\_2496.erunat\_14062003\_175645\_16271 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:\*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*

29: em.vi:\*  
30: em.htg\_hum:\*  
31: em.htg\_inv:\*  
32: em.htg\_other:\*  
33: em.htg\_mus:\*  
34: em.htg\_pln:\*  
35: em.htg\_rod:\*  
36: em.htg\_mam:\*  
37: em.htg\_vrt:\*  
38: em.sy:\*  
39: em.htgo\_hum:\*  
40: em.htgo\_mus:\*  
41: em.htgo\_other:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1086	100.0	596	9	AF529823 Homo sapi
2	1086	100.0	596	9	AF529826 Homo sapi
3	1086	100.0	596	12	AF529837 Mus muscu
4	1086	100.0	1837	9	BC006296 Homo sapi
5	1086	100.0	2791	9	AB040431 Homo sapi
6	1084	99.8	596	9	AF529815 Homo sapi
7	1083	99.7	596	12	AF529833 Mus muscu
8	1083	99.7	596	12	AF529835 Mus muscu
9	1082	99.6	596	12	AF529830 Mus muscu
10	1082	99.6	596	12	AF529831 Mus muscu
11	1082	99.6	596	12	AF529838 Mus muscu
12	1081	99.5	596	9	AF529819 Homo sapi
13	1081	99.5	596	9	AF529822 Homo sapi
14	1081	99.5	596	9	AF529825 Homo sapi
15	1081	99.5	596	12	AF529829 Mus muscu
16	1081	99.5	596	12	AF529832 Mus muscu
17	1080	99.4	595	12	AF529834 Mus muscu
18	1080	99.4	596	9	AF529827 Homo sapi
19	1079	99.4	596	9	AF529820 Homo sapi
20	1079	99.4	596	9	AF529821 Homo sapi
21	1079	99.4	597	12	AF529828 Mus muscu
22	1078	99.3	596	12	AF529836 Mus muscu
23	1078	99.3	596	12	AF529839 Mus muscu
24	1076	99.1	597	12	AF529842 Cricetulu
25	1076	99.1	597	12	AF529843 Cricetulu
26	1076	99.1	597	12	AF529844 Cricetulu
27	1076	99.1	597	12	AF529847 Cricetulu
28	1076	99.1	597	12	AF529852 Cricetulu
29	1075	99.0	596	9	AF529824 Homo sapi
30	1072	98.7	596	9	AF529816 Homo sapi
31	1072	98.7	596	9	AF529818 Homo sapi
32	1072	98.7	597	12	AF529845 Cricetulu
33	1071	98.6	597	12	AF529851 Cricetulu
34	1070	98.5	596	9	AF529817 Homo sapi
35	1070	98.5	597	12	AF529841 Cricetulu
36	1070	98.5	597	12	AF529846 Cricetulu
37	1069	98.4	597	12	AF529848 Cricetulu
38	1069	98.4	597	12	AF529849 Cricetulu
39	1067	98.3	597	12	AF529850 Cricetulu
40	1065	98.1	597	12	AF529853 Cricetulu
41	1065	98.1	597	12	AF529854 Cricetulu
42	1057	97.3	591	12	AF529855 Cricetulu
43	1044	96.1	577	12	AF529840 Mus muscu
44	1008	92.8	2440	10	AF132979 Mus muscu
45	986	90.8	547	12	AF529856 Cricetulu

ALIGNMENTS

RESULT 1

AF529823	AF529823	596 bp	linear	PRI 19-AUG-2002
LOCUS	Homo sapiens clone Ramos 9 AID (AID)	mrna		
DEFINITION	Homo sapiens clone Ramos 9 AID (AID)	mrna	partial cds.	
ACCESSION	AF529823			
VERSION	AF529823.1	GI:22297233		
KEYWORDS	human.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
AUTHORS	1 (bases 1 to 596)			
TITLE	Martin, A. and Scharff, M.D.			
JOURNAL	Somatic hypermutation of the AID transgene in B cells and non-B cells			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 596)			
TITLE	Martin, A. and Scharff, M.D.			
JOURNAL	Direct Submission			
FEATURES	Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA			
source	Location/Qualifiers			
	1..596			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
gene	1..>596			
	/gene="AID"			
CDS	1..>596			
	/gene="AID"			
	/note="integrated into Burkitt's lymphoma cell line Ramos"			
	/codon_start=1			
	/product="Aid"			
	/protein_id="AAM95410.1"			
	/db_xref="GI:22297234"			
	/translation="MDSLNNRRKFLYQFNVRWAKGRRETYLCYVVKRDSATSFSLDFGLRNKNGCHEVLEFLRYISDWLDLPGRCYRTWTFSSPCYDCARHVADFLRGNP			
	NLSRLIFATRLYCEQDKAPGELRLHRAVGQIAIMTFNDYFCWNTFVENHRTFFK			
	AWELGHENSRYLSRQRLRILLPLYEVDLLDAFRTLGL"			
BASE COUNT	128 a 163 c 155 g 150 t			
ORIGIN				
Alignment Scores:				
Pred. No.:	3,81e-109	Length:	596	
Score:	1086.00	Matches:	198	
Percent Similarity:	100.00%	Conservative:	0	
Best Local Similarity:	100.00%	Mismatches:	0	
Query Match:	100.00%	Indels:	0	
DB:	9	Gaps:	0	
US-09-966-880A-8 (1-198) x AF529823 (1-596)				
Qy	1 MetAspSerLeuLeuMetAsnArgArgGlyPheLeuTyrGlnPheLysAsnValArgTtp	20		
Db	1 ATGCACAGCCATATGATGAACCCGAGAGAGATTCTTTACCAATTCAAAATGTCCTGG	60		
Qy	21 AlAlYsGlyArgArgGluThrTyrLeuCyTyrValValLysArgArgAspSerAlaThr	40		
Db	61 GCTAAGGGTCGGCGTGAGACCTACCTGCTAGTAGTGAAGAGCGGTGACATGGCTACA	120		
Qy	41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu	60		
Db	121 TCCTTTTCACTGACATTGGTTATCTTCGCAATPAAGACGGCTGCCACGTGGAAATTCCTC	180		
Qy	61 PheLeuArgTyrIleSerAspTrrAspLeuAspProGlyArgCysTyrArgValThrTrr	80		
Db	181 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACTGG	240		
Qy	81 PheThrSerTrrSerProCyTrrAspCysAlaArgHisValAlaAspPheLeuArgGly	100		
Db	241 TTCACCTCTCTGGAGCCCTGCTACGACTGTCGCCGACATGTGCCGACATTTCTGCCAGGG	300		
Qy	101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTrrPheCysGluAspArgIlys	120		

```

QY 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGCGTGGCGTGGACCTACCTGTGCTACGTAGTGAAGCGGTGACAGTCTACA 120
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60
Db 121 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAACGGGTGCACAGTGGAAATGCTC 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCCTCCCGCTACATCTCGAGCTGGACCTAGACCTGGCGGTGCTACCGGTCACTCCG 240
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCCGACATGTGGCGGACTTCTTCGGAGGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AATCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCGAGGGCTGGCGGTGCACCGCGGGGTGCAAAATAGCCATCATGACC 420
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTGTGAGAAACCATGAAGAACTTTCAAA 480
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgIleLeu 180
Db 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCTT 540
QY 181 LeuProLeuTyrGluValAspLeuAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCATTTCGTACTTTGGGACTT 594

RESULT 3
AF529837
LOCUS
DEFINITION Mus musculus clone 10 transgenic Homo sapiens AID (AID) mRNA,
complete cds.
ACCESSION AF529837
VERSION AF529837.1 GI:22297261
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B and non-B cells
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Direct Submission
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
    location/Qualifiers
    1..596
    /organism="Mus musculus"
    /db_xref="taxon:10090"
    /clone="10"
    /cell_type="hybridoma P1-5"
    /transgenic
    1..596
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    1..>596
    /gene="AID"
    1..>596
    /gene="AID"
    /codon_start=1

source
    BC006296
    1837 bp mRNA linear PRI 12-JUL-2001
    Homo sapiens, activation-induced cytidine deaminase, clone
    MGC:12911 IMAGE:4054915, mRNA, complete cds.
    ACCESSION BC006296
    VERSION BC006296.1 GI:13623400
    KEYWORDS MGC.
    SOURCE Homo sapiens.
    ORGANISM Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
    REFERENCE 1 (bases 1 to 1837)

```

```

/product="AID"
/protein_id="AAM95424.1"
/db_xref="GI:22297262"
/translation="MDSLMMRRKFLYQFKVNRWAKGRRETYLCYVVRDSDATSFSL
NFGYLNRNKGCHVELELLRYISDWDLPGRCYRVTFWFTSMSPCYDCARRVADFLRGNP
DLGRIETFARLYFCEDRKAEPGLRHRAGVQIAIMTFKDYFCYCNWTFVENHETFK
AWEGIHENSVRSLRQLRILLPLLYEVDLDRAFTLGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN
Alignment Scores:
  3.81e-109 Length: 596
  Score: 1086.00 Matches: 198
  Percent Similarity: 100.00% Conservative: 0
  Best Local Similarity: 100.00% Mismatches: 0
  Query Match: 100.00% Indels: 0
  Db: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529837 (1-596)
QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGGACACCCCTTGTGTAACCGGAGGAGTTTCTTTACCAATTCAAAATGTCCGCTGG 60
QY 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGCGTGGCGTGGACCTACCTGTGCTACGTAGTGAAGCGGTGACAGTCTACA 120
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 60
Db 121 TCCTTTTTCACGTGGACTTTGGTTATCTTCGCAATAAGAACGGGTGCACAGTGGCTC 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCCTCCCGCTACATCTCGAGCTGGACCTAGACCTGGCGGTGCTACCGGTCACTCG 240
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCCGACATGTGGCGGACTTCTTCGGAGGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCGAGGGCTGGCGGTGCACCGCGGGGTGCAAAATAGCCATCATGACT 420
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTCAAA 480
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCTT 540
QY 181 LeuProLeuTyrGluValAspLeuAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 541 TTGCCCTGTATGAGTTGATGACTTACGAGACGCATTTCGTACTTTGGGACTT 594

RESULT 4
BC006296
LOCUS
DEFINITION Homo sapiens, activation-induced cytidine deaminase, clone
MGC:12911 IMAGE:4054915, mRNA, complete cds.
ACCESSION BC006296
VERSION BC006296.1 GI:13623400
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1837)

```

# AUTHORS TITLE JOURNAL

Strausberg, R.  
Direct Submission  
Submitted (09-APR-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
Tissue Procurement: Louis Staudt  
cDNA Library Preparation: Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nigri.nih.gov](mailto:nisc_mgc@nigri.nih.gov)  
Shvchenko, Y., Wethersby, K.D., Beckstrom-Sternberg, S.M.,  
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
Lim, M., Maduro, Q.L., Masiello, C., Mastrian, S.D., McCloskey, J.C.,  
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,  
Tongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A.,  
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAL Plate: 17 Row: a Column: 1  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 9988409.

## FEATURES

### source

1. .1837  
/organism="Homo sapiens"  
/db\_xref="LocusID:57379"  
/clone="MGC:12911 IMAGE:4054915"  
/tissue\_type="Primary B-Cells from Tonsils"  
/clone\_lib="NIH\_MGC\_48"  
/lab\_host="DH10B-R"  
/notes="Vector: pOT87"  
86. .682  
/codon\_start=1  
/product="activation-induced cytidine deaminase"  
/protein\_id="AAH06296.1"  
/db\_xref="GI:13623401"  
/translation="MDSLLMNRKELYQFNVRWAKGRRETYLVVVKRDSATSFSL  
DFGLRNKNGCHELLEFLRYISDMDLDPGRCYRVWFTWSFCYDCARHVAFLRNP  
NLSRLFTARLYFCEDRKAEPGLRLRHAGVOIALMTFDYFYCWNTEVFNHERTFK  
AWEGLEHNSVRLSRQLRILLPLYEVDLLRDAFRTLGL"

### CDS

BASE COUNT 530 a 387 c 421 g 499 t  
ORIGIN

## Alignment Scores:

Pred. No.: 1.55e-108 Length: 1837  
Score: 1086.00 Matches: 198  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x BC006296 (1-1837)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrglnPheLysAsnValArgTrrp 20  
|||||  
DB 86 ATGCAGACCTCTGTGATGACCGAGGAGGAGTTCTTACCAATTCAAAATGTCGCTGG 145  
QY 21 AlaLysGlyArgArgGluThrTyrrLeuCystrTyrrValValLysArgArgSsrAlaThr 40  
|||||  
DB 146 GCTAAGGGTGGCGGTGAGACCTACCTGTGTCTAGCTAGTGAAGAGCGGTGACGTGTACA 205  
QY 41 SerPheSerLeuAspPheGlyTyrrLeuArgAsnLysAsnGlyCysHisValGluLeu 60  
|||||  
DB 206 TCCTTTTCTACTGGACTTGTGGTTATCTTCGCAATAAGAACGGCTGCCACGTTGGAATTGCTC 265

QY 61 PheLeuArgTyrrIleSerAspTrpAspLeuAspProGlyArgCysTyrrArgValThrTrp 80  
|||||  
DB 266 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGTACCGCTGACCTGG 325  
QY 81 PheThrSerTrpSerProCysTyrrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
|||||  
DB 326 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCCAGGG 385  
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrrPheCysGluAspArgLys 120  
|||||  
DB 386 AACCCCAACCTCAGTCTGAGGACTTTCACCGCGCGCTTACTTCTGTGAGGACCCCAAG 445  
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
|||||  
DB 446 GCTGAGCCCGAGGGGCTGCGGGGCTGCACCGCGCGGGTGCATAATGCCATCATGACC 505  
QY 141 PheLysAspTyrrPheTyrrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
|||||  
DB 506 TTCAAAGATTATTTTACTGCTGGATACTTTTGTAGAAACCATGAAAGAACTTTTCAA 565  
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgGlyLeu 180  
|||||  
DB 566 GCTGTGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCCATCCCTT 625  
QY 181 LeuProLeuTyrrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198  
|||||  
DB 626 TTGCCCTGTATGAGGTTGATGACTTACGAGAGGCATTTTCGTACTTTGGGACTT 679  
RESULT 5  
LOCUS AB040431 2791 bp mRNA linear PRI 03-OCT-2000  
DEFINITION Homo sapiens AID mRNA for activation-induced cytidine deaminase,  
complete CDS.  
ACCESSION AB040431  
VERSION AB040431.1 GI:9988409  
KEYWORDS AID; activation-induced cytidine deaminase; Human AID.  
SOURCE Homo sapiens cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (sites)  
Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.  
Isolation, tissue distribution, and chromosomal localization of the  
human activation-induced cytidine deaminase (AID) gene  
Genomics 68 (1), 85-88 (2000)  
20408890  
2 (sites)  
Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O.,  
Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A.,  
Tezcan, I., Ersoy, F., Kayserili, H., Ugazio, A.G., Brousse, N.,  
Muramatsu, M., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A.  
and Durandy, A.  
Activation-induced cytidine deaminase (AID) deficiency causes the  
autosomal recessive form of the Hyper-IgM syndrome (HIGM2)  
Cell 102 (5), 565-575 (2000)  
20460541  
3 (bases 1 to 2791)  
Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.  
Direct Submission  
Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department  
of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku,  
Kyoto, Kyoto 606-8501, Japan (E-mail: honjo@four.med.kyoto-u.ac.jp,  
Tel: 81-75-753-4371(ex. 4371), Fax: 81-75-753-4388)  
FEATURES  
source Location/Qualifiers  
1. .2791  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1. .2791  
/gene="AID"  
77. 673  
/gene="AID"  
/codon\_start=1



/product="activation-induced cytidine deaminase"  
/protein\_id="BAB12721.1"  
/db\_xref="GI:9988410"  
/translation="MDSLMMNRKFLYQKVRWAKGRRETYLCYVVKRRDSATSFSL  
DFGYLRNKGCHVELLFURYISDWDLDFGRCYRVTWFTSWSPCYDCARHVAFLRGNP  
NLSLRIFARLYFCEDRAEPEGLRKHAGVQIAITMFKDYFCWNTEVENHERTEFK  
AWEGUHENSVRLSRQIRILLPLXYEVDLDRDAFRTLGL"  
BASE COUNT 842 a 548 c 625 g 776 t  
ORIGIN

Alignment Scores:  
Pred. No.: 2,61e-108 Length: 2791  
Score: 1086.00 Matches: 198  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 100.00% Indels: 0  
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AB040431 (1-2791)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
DB 77 ATGGACACCTCTTGATGACCGGAGGAGTTCTTTACCAATTCAAAAATGTCGCGTGG 136  
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
DB 137 GCTAAGGTCGGCGTGAGACCTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTCTACA 196  
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
DB 197 TCCTTTTCACTGGAGCTTGGTTATCTTCGCAATAAGACGGCTGCCAGTGGAAATGTGCTC 256  
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
DB 257 TTCTCCCGCTACATCTCGAGCTGGGACCTAGACCTGCGCGTGTACCGCGTCACTGG 316  
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
DB 317 TTCACCTCTGGAGCCCTGCTGACGACTGTGCCGACATGTGGCGACTTTCGCGAGGG 376  
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
DB 377 AACCCCAACCTCAGTCTGAGGATCTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 436  
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
DB 437 GCTGAGCCGAGGGCTCGCGCGCTGACCCGCGGGGTGCAAAATAGCATATGACC 496  
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
DB 497 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 556  
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgIleLeu 180  
DB 557 GCCTGGGAAGGCTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTGGCGCATCTT 616  
QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198  
DB 617 TTGCCCCGTATGAGTTGATGACTTACGAGAGCGCATTTTCGTACTTTGGGACTT 670

RESULT 6  
AF529815  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 596)  
AUTHORS  
Martin.A. and Scharff,M.D.

.TITLE Somatic hypermutation of the AID transgene in B cells and non-B cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin.A. and Scharff,M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
Location/Qualifiers  
1..596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Ramos 1"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/note="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95402.1"  
/db\_xref="GI:22297218"  
/translation="MDSLMMNRKFLYQKVRWAKGRRETYLCYVVKRRDSATSFSL  
DFGYLRNKGCHVELLFURYISDWDLDFGRCYRVTWFTSWSPCYDCARHVAFLRGNP  
NLSLRIFARLYFCEDRAEPEGLRKHAGVQIAITMFKDYFCWNTEVENHERTEFK  
AWEGUHENSVRLSRQIRILLPLXYEVDLDRDAFRTLGL"  
BASE COUNT 128 a 163 c 155 g 150 t  
ORIGIN

Alignment Scores:  
Pred. No.: 6,3e-109 Length: 596  
Score: 1084.00 Matches: 197  
Percent Similarity: 100.00% Conservative: 1  
Best Local Similarity: 99.49% Mismatches: 0  
Query Match: 99.82% Indels: 0  
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AF529815 (1-596)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
DB 1 ATGGACACCTCTTGATGACCGGAGGAGTTCTTTACCAATTCAAAAATGTCGCGTGG 60  
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
DB 61 GCTAAGGTCGGCGTGAGACCTACCTGTCTACGTAGTCAAGAGCGGTGACAGTGTCTACA 120  
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
DB 121 TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACCTGGAATGTGCTC 180  
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
DB 181 TTCTCCCGCTACATCTCGGACTGGGACCTAGACCTTGGCGCTGTCTACCGGTCACTGG 240  
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
DB 241 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTCTCGCGAGGG 300  
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
DB 301 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 360  
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
DB 361 GCTGAGCCCGAGGGGCTGCGCGGTGTCACCGCGGGGTGCAAAATAGCATATGACC 420  
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
DB 421 TTCAAAGATTATTTTACTGCTGGAAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA 480  
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgIleLeu 180  
DB 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgIleLeu 180

Db	481	GCCTGGGAGGCTGCATGAAATTCAGTTCCTCTCTCCAGACAGATTCCGGCGCATCCTT	540
Qy	181	LeuProLeuTyrrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu	198
Db	541	TTGCCCTGTATGAGTTGATGACTTACGACAGCGATTTCGTACTTTGGGACTT	594
RESULT 7			
AF529833			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
JOURNAL			
REFERENCE			
AUTHORS			
JOURNAL			
FEATURES			
source			
source			
gene			
CDS			
BASE COUNT	127 a	163 c	156 g 150 t
ORIGIN			
Alignment Scores:			
Pred. No.:	8.09e-109	Length:	596
Score:	1083.00	Matches:	197
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.49%	Mismatches:	0
Query Match:	99.72%	Indels:	0
DB:	12	Gaps:	0
US-09-966-880A-8 (1-198) x AF529833 (1-596)			
Qy	1	MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrrGlnPheLysAsnValArgTrp	20
Db	1	ATGGACAGCCTCTTGATGAACCGGAGGAGTTCCTTACCAATTCAAAAATGTCGCGTG	60
Qy	21	AlaLysGlyArgArgGluThrTyrrLeuCysTyrrValValLysArgArgAspSerAlaThr	40
Db	61	GCTAAGGGTCGCGTGAGACCTACGTTGCTACGTAGTAGCAAGGCGTGACAGTGCCTAC	120
Qy	41	SerPheSerLeuAspPheGlyTyrrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu	60
Db	121	TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGACGCGTGCACGCGAATTGCTC	180
Qy	61	PheLeuArgTyrrIleSerAspTrpAspLeuAspProGlyArgCysTyrrArgValThrTrp	80
Db	181	TTCTTCGCGTACATCTCGGACTGGGACCTAGACCTGCGCGCTGCTACCGCGTCACCTGG	240
Qy	81	PheThrSerTrpSerProCysTyrrAspCysAlaArgHisValAlaAAspPheLeuArgGly	100
Db	241	TTACCTCTCGGAGCCCTGCTACGACTGTGCCCGACATGTGGCGGACTTTCTCGGAGGG	300
Qy	101	AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrrPheCysGluAspArgLys	120
Db	301	AACCCCAACCTCAGTCTCAGGATCTTCCCGCGCGCTCTACTTCTGTGAGGACCGCAAG	360
Qy	121	AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr	140
Db	361	GCTGAGCCCGAGGGCTGCGCGGCTGCACGCGCGGGTGCATAATAGCATCATGACC	420
Qy	141	PheLysAspTyrrPheTyrrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys	160
Db	421	TTCAAGATTATTTTACTGCTGGAATACITTTGTAGAAAACCATGAAGAATTTTCAAA	480
Qy	161	AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu	180
Db	481	GCCTGGGAGGCTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGCGCATCCTT	540
Qy	181	LeuProLeuTyrrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu	198
Db	541	TTGCCCTGTATGAGTTGATGACTTACGACAGCGCATTCGTACTTTGGGACTT	594
RESULT 8			
AF529835			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
JOURNAL			
REFERENCE			
AUTHORS			
JOURNAL			
FEATURES			
source			
source			
gene			
CDS			
BASE COUNT	127 a	163 c	156 g 150 t
ORIGIN			
Alignment Scores:			
Pred. No.:	8.09e-109	Length:	596
Score:	1083.00	Matches:	197
Percent Similarity:	100.00%	Conservative:	1
Best Local Similarity:	99.49%	Mismatches:	0
Query Match:	99.72%	Indels:	0
DB:	12	Gaps:	0
US-09-966-880A-8 (1-198) x AF529833 (1-596)			
Qy	1	MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrrGlnPheLysAsnValArgTrp	20
Db	1	ATGGACAGCCTCTTGATGAACCGGAGGAGTTCCTTACCAATTCAAAAATGTCGCGTG	60
Qy	21	AlaLysGlyArgArgGluThrTyrrLeuCysTyrrValValLysArgArgAspSerAlaThr	40
Db	61	GCTAAGGGTCGCGTGAGACCTACGTTGCTACGTAGTAGCAAGGCGTGACAGTGCCTAC	120
Qy	41	SerPheSerLeuAspPheGlyTyrrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu	60
Db	121	TCCTTTTCTACTGGACTTTGGTTATCTTCGCAATAAGACGCGTGCACGCGAATTGCTC	180

ORIGIN		FEATURES		Location/Qualifiers	
Alignment Scores:		Source		1..596	
Pred. No.:	8-09e-109	Length:	596	/organism="Mus musculus"	
Score:	1083.00	Matches:	197	/db_xref="taxon:10090"	
Percent Similarity:	100.00%	Conservative:	1	/clone="3"	
Best Local Similarity:	99.49%	Mismatches:	0	/cell_type="hybridoma Pl-5"	
Query Match:	99.72%	Indels:	0	/transgenic	
DB:	12	Gaps:	0	1..596	
US-09-966-880A-8 (1-198) x AF529835 (1-596)		gene		/organism="Homo sapiens"	
		CDS		/db_xref="taxon:9606"	
				1..596	
				/gene="AID"	
				/codon_start=1	
				/product="AID"	
				/protein_id="AA95417.1"	
				/db_xref="GI:22297248"	
				/translation="MDSLIMNRKKFLYQFNVRWAKGRRETYLCYVVVKKRDSATSFSL DFGLRNKNGCHVELLFLRYISDWLDPGRCYRVTFWTFSPCYDCARHVADFRLNRP NLSLRFTARLYFCEDRKAEPGLRRLRHAGVQIAIMSFKDYFCWNTFVENHRTFFK AMEGLHENSVRLSROLRRILLPLYEVDLLRDAFTLGL"	
BASE COUNT	126 a	164 c	155 g	151 t	
ORIGIN					
Alignment Scores:		1-04e-108		Length: 596	
Pred. No.:	US-09-966-880A-8 (1-198) x AF529830 (1-596)	Score:		1082.00	
Score:		Percent Similarity:		100.00%	
Percent Similarity:		Best Local Similarity:		99.49%	
Best Local Similarity:		Query Match:		99.63%	
DB:		Indels:		0	
DB:		Gaps:		0	
US-09-966-880A-8 (1-198) x AF529830 (1-596)					
QY	1	MetAspSerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrrp	20		
Db	1	ATGGACAGCCTCTGATGAACGGAGGAAGTTCTTTTACCAATTCAAAAATGTCGCGTGG	60		
QY	21	AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr	40		
Db	61	GCTAAGGGTGGCGTGAGACCTACCTGTGCTAGTAGAAGAGCGGTGACAGTGTCTACA	120		
QY	41	SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu	60		
Db	121	TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGGATATGCTC	180		
QY	61	PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp	80		
Db	181	TTCTCTCCGCTACATCTCGGACTTGGGACCTAGACCTTGGCGCTGCTACCGCGTCACCTGG	240		
QY	81	PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly	100		
Db	241	TTACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCCGGCTTCTTCCGAGGG	300		
QY	101	AsnProAsnLeuSerLeuArgGilePheThrAlaArgLeuTyrPheCysGluAspArgLys	120		
Db	301	AAACCCCAACTCAGTCTGAGGATCTTACCGCGCGCTCTACTCTGTGAGGACCGCAAG	360		
QY	121	AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr	140		
Db	361	GCTGAGCCCGAGGGGCTGCGGGCTGCACCGCGCGGGTGCACATAGCATCATGTCC	420		
QY	141	PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys	160		
Db	421	TTCAAGATATATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAAGAACTTTCAA	480		
QY	161	AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu	180		
Db	481	GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCCGCGCATCTT	540		
QY	181	LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu	198		
Db	541	TTGCCCTGTATGAGGTTGATGACTTACGAGACGCAATTTCTGTTGGGACTT	594		
RESULT 9					
AF529830					
LOCUS					
DEFINITION					
complete cds.					
ACCESSION					
AF529830					
VERSION					
AF529830.1					
KEYWORDS					
house mouse.					
SOURCE					
Mus musculus					
ORGANISM					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					

```

Db      541 TTGCCCCCTGATGAGGTTGATGACTTACGAGAGCGCATTTTCGTACTTTGGGACTT 594

RESULT 10
AF529831
LOCUS   Mus musculus clone 4 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529831
VERSION   AF529831.1 GI:22297249
KEYWORDS
SOURCE   house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
JOURNAL Somatic hypermutation of the AID transgene in B and non-B cells
TITLE     2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
JOURNAL Direct Submission
SUBMITTED (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
source
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="4"
/cell_type="hybridoma P1-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95418.1"
/db_xref="GI:22297250"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFLRYISDWLDPGRCYRVWFTSPCYDCARHVDLRCNPN
NLSRIFTARLYFCEDRKAEPGLRLRHAGVOIVMTPKDYFCWNTFVENHERTFK
AWEGHNSVRLSRQLRLLPLYEVDLDRDARTLGL"
BASE COUNT 127 a 163 c 155 g 151 t
ORIGIN
Alignment Scores:
Pred. No.: 1.04e-108 Length: 596
Score: 1082.00 Matches: 197
Percent Similarity: 99.49% Conservative: 0
Best Local Similarity: 99.49% Mismatches: 1
Query Match: 99.63% Indels: 0
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529831 (1-596)

QY      1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrrp 20
Db      1 ATGGACAGCCTCTTGATGAACCGGAGGAGGAGTTCTTTACCAATTCAAAAATGTCGCGTGG 60
QY      21 AlaLysGlyArgArgGluThrTyrLeuCystTyrValValLysArgArgAspSerAlaThr 40
Db      61 GCTAAGGGTCGCGGTGAGACCTACCTGCTACGTAGTGAAGAGCGGTGACAGTGCTACA 120
QY      41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db      121 TCCTTTTCACTGAGACTTTGGTTATCTTCGCAATAGAACGGCTGCCAGTGGAAATTGCTC 180
QY      61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrrp 80
Db      181 TTCCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCCTGCTACCGCGTCACTGG 240

```

```

QY      81 PheThrSerTrrpSerProCysTyrAspCysAlaArgHisValAlaAAspPheLeuArgGly 100
Db      241 TTCACCTCCTGGAGCCCTGCTACGACTGTGCCCGACATGTGGCGGACTTTTCTCGGAGGG 300
QY      101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db      301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCCTCTACTTCTGTGAGGACCGCAAG 360
QY      121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db      361 GCTGAGCCGCGGCGGCTGCCGCGCTCACCGCGCGGGGTGCAATAGGCATCATGACC 420
QY      141 PheLysAspTyrPheTyrCysTrrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db      421 TTCAAGATATTATTTTACTGCTGGAATACCTTTGTAGAAAAACCATGAAGAACTTTTCAA 480
QY      161 AlaTrrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db      481 GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGTTTCGCGCATCCTT 540
QY      181 LeuProLeuTrrpGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db      541 TTGCCCTGTATGAGTTGATGACTTACGAGAGCGCATTTTCGTACTTTGGGACTT 594

```

```

RESULT 11
AF529838
LOCUS   Mus musculus clone 11 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529838
VERSION   AF529838.1 GI:22297263
KEYWORDS
SOURCE   house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
JOURNAL Somatic hypermutation of the AID transgene in B and non-B cells
TITLE     2 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
JOURNAL Direct Submission
SUBMITTED (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES
Location/Qualifiers
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="11"
/cell_type="hybridoma P1-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AAM95425.1"
/db_xref="GI:22297264"
/translation="MDSLMMNRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSL
DFGLRNKNGCHVELLFLRYISDWLDPGRCYRVWFTSPCYDCARHVDLRCNPN
NLSRIFTARLYFCEDRKAEPGLRLRHAGVOIVMTPKDYFCWNTFVENHERTFK
AWEGHNSVRLSRQLRLLPLYEVDLDRDARTLGL"
BASE COUNT 128 a 163 c 154 g 151 t
ORIGIN

```

Alignment Scores: 1.04e-108 Length: 596  
Pred. No.:

```
Score: 1082.00 Matches: 197
Percent Similarity: 99.49% Conservative: 0
Best Local Similarity: 99.49% Mismatches: 1
Query Match: 99.63% Indels: 0
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x AF529838 (1-596)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrrp 20
Db 1 ATGGACACCTCTTGATGACCGGAGGAGTCTTTTACCAATTCACCAATTCGCGTGG 60

Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
Db 61 GCTAAGGTCGGCGTGAGACCTACCTGCGCGCTGCTACCGGCTACCTGG 120

Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCTTTTTCACCTGGAGCTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCGCTC 180

Qy 61 PheLeuArgTyrIleSerAspTyrAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCCTCCGCTACATCTCGACCTGAGACCTAGACCTGCGCGCTGCTACCGGCTACCTGG 240

Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCCTGGAGCCCTGCTACGACCTGTGCCGACATGTGCCGACCTTTTCACAA 360

Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCCACCGCGCCCTCTACTTCTGTGAGGACCGCAAG 360

Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTAGCCCGAGGGCTCGCGGCTGACCGCGCGGGGTACAAATAGTCATCATGACC 420

Qy 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAGATATATTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAGAATTTTCAAA 480

Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGGCGCATCTT 540

Qy 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 541 TTGCCCCGTATGAGGTGATGACTTACGACACGCAATTCGTACTTTGGGACTT 594

RESULT 12
AF529819 596 bp mRNA linear PRI 19-AUG-2002
LOCUS Homo sapiens clone Ramos 5 AID (AID) mRNA, partial cds.
DEFINITION AF529819
ACCESSION AF529819.1 GI:22297225
KEYWORDS human.
SOURCE
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 596)
AUTHORS Martin,A. and Scharff,M.D.
TITLE Somatic hypermutation of the AID transgene in B cells and non-B
cells
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 596)
TITLE Martin,A. and Scharff,M.D.
JOURNAL Direct Submission
SUBMITTED (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
FEATURES Location/Qualifiers
Source 1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"

/clone="Ramos 5"
1..>596
/gene="AID"
1..>596
/gene="AID"
/note="Integrated into Burkitt's lymphoma cell line Ramos"
/codon_start=1
/product="AID"
/protein_id="AAM95406.1"
/db_xref="GI:22297226"
/translation="MDSLMMNRKFLYQKNNRWAKGRRETYLCVIVVRRDSATSFSL
DFGLRNKNGCHVELLFYISDWLDPRGRCYRTWFTSWSPCYDCARHVAFLRGNP
NLSLRFIATRYFCEDHKAEPGLRHLRAGVQIAITMFKDYCWNTPEVHERHTFK
AWEGILHNSVRLSRQLRILLPLYEVDLDRAFRTLGL"
BASE COUNT 128 a 164 c 154 g 150 t
ORIGIN

Alignment Scores:
Pred. No.: 1.34e-108 Length: 596
Score: 1081.00 Matches: 197
Percent Similarity: 99.49% Conservative: 0
Best Local Similarity: 99.49% Mismatches: 1
Query Match: 99.54% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AF529819 (1-596)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrrp 20
Db 1 ATGGACACCTCTTGATGACCGGAGGAGTCTTTTACCAATTCACCAATTCGCGTGG 60

Qy 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr 40
Db 61 GCTAAGGTCGGCGTGAGACCTACCTGCGCGCTGCTACCGGCTACCTGGTACA 120

Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCTTTTTCACCTGGAGCTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTCGCTC 180

Qy 61 PheLeuArgTyrIleSerAspTyrAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCCTCCGCTACATCTCGACCTGAGACCTAGACCTGCGCGCTGCTACCGGCTACCTGG 240

Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCCTGGAGCCCTGCTACGACCTGTGCCGACATGTGCCGACCTTTTCACAA 360

Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGGATCTTCCACCGCGCCCTCTACTTCTGTGAGGACCGCAAG 360

Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTAGCCCGAGGGCTCGCGGCTGACCGCGCGGGGTACAAATAGTCATCATGACC 420

Qy 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAGATATATTTTACTGCTGGAATCTTTTGTAGAAAACCATGAAGAATTTTCAAA 480

Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 481 GCCTGGGAAGGGCTGCATGAAATTCAGTTCTGCTCTCCAGACAGCTTCGGCGCATCTT 540

Qy 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 541 TTGCCCCGTATGAGGTGATGACTTACGACACGCAATTCGTACTTTGGGACTT 594

RESULT 13
AF529822 596 bp mRNA linear PRI 19-AUG-2002
LOCUS Homo sapiens clone Ramos 8 AID (AID) mRNA, partial cds.
DEFINITION AF529822
ACCESSION AF529822
VERSION AF529822.1 GI:22297231
```

KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE 1 (bases 1 to 596)  
Somatic hypermutation of the AID transgene in B cells and non-B  
cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source  
Location/Qualifiers  
1..596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Ramos 8"  
1..>596  
/gene="AID"  
1..>596  
/notes="Integrated into Burkitt's lymphoma cell line Ramos"  
/product="Aid"  
/protein\_id="AAM95409.1"  
/db\_xref="GI:22297232"  
/translation="MDSLMMNRKFLYQFNVRWAKGRREYLCYVVKRRDRTSFL  
DFGLRNKNGCHVELLFLRYISDWLDPGRCYRVWFTSWSPCYDCARHVAFLRGNP  
NLSLRIFTALRYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHRTFK  
AWEGLEHNSVRLSRQLLLPLYEVDLDRFAFTLGL"  
BASE COUNT 126 a 165 c 155 g 150 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1 34e-108 Length: 596  
Score: 1081.00 Matches: 197  
Percent Similarity: 99.49% Conservative: 0  
Best Local Similarity: 99.49% Mismatches: 1  
Query Match: 99.54% Indels: 0  
DB: 9 Gaps: 0  
US-09-966-880A-8 (1-198) x AF529822 (1-596)  
QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 60  
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
Db 61 GCTAAGGGTCGGGTGAGACCTACCTGTGCTACGTAGTAGAAGAGCGGTGACCGTGTCTACA 120  
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
Db 121 TCCCTTTCAGTGGCTTGGTTATCTTCGCATAGACGGCTGCCACGTGGAATGCTC 180  
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
Db 181 TTCCCTCCGCTACATCTCGSACTGGGACCTAGACCCCTGCGCGCTGCTACCGCTCACCTGG 240  
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
Db 241 TTCACCTCTGGAGCCCTCTGCTACGACGTGCGCCGACATGTGGCGACTTTCTGCGAGGG 300  
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
Db 301 AACCCCAACCTCAGTCTGAGGATCTCACCGCGCGCCCTACTTCTGTGTGAGACCGCAG 360  
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMethThr 140  
Db 361 GCTGAGCCCGAGGGGCTCGCGCGCTGACCGCGCGGGGTGCAAAATAGCCATCATGACC 420

QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
Db 421 TTCAAGATTTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAACTTTCAAA 480  
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgGlyLeu 180  
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCACAGACGCTTCGGCGCATCCTT 540  
QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198  
Db 541 TTGCCCTGTATGAGTGTGATGACTTACGAGAGCGCATTTCTGCTACTTTGGGACTT 594  
RESULT 14  
AF529825  
LOCUS Homo sapiens clone Ramos 11 AID (AID) mRNA, partial cds.  
DEFINITION AF529825  
ACCESSION AF529825  
VERSION AF529825.1 GI:22297237  
KEYWORDS human.  
SOURCE  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE 1 (bases 1 to 596)  
Somatic hypermutation of the AID transgene in B cells and non-B  
cells  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 596)  
AUTHORS Martin, A. and Scharff, M.D.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of  
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA  
FEATURES  
source  
Location/Qualifiers  
1..596  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Ramos 11"  
1..>596  
/gene="AID"  
1..>596  
/gene="AID"  
/notes="Integrated into Burkitt's lymphoma cell line Ramos"  
/codon\_start=1  
/product="AID"  
/protein\_id="AAM95412.1"  
/db\_xref="GI:22297238"  
/translation="MDSLMMNRKFLYQFNVRWAKGRREAYLCYVVKRRDSATSFSL  
DFGLRNKNGCHVELLFLRYISDWLDPGRCYRVWFTSWSPCYDCARHVAFLRGNP  
NLSLRIFTALRYFCEDRKAEPGLRLRHAGVQIAIMTFKDYFCWNTFVENHRTFK  
AWEGLEHNSVRLSRQLLLPLYEVDLDRFAFTLGL"  
BASE COUNT 126 a 164 c 156 g 150 t  
ORIGIN  
Alignment Scores:  
Pred. No.: 1 34e-108 Length: 596  
Score: 1081.00 Matches: 197  
Percent Similarity: 99.49% Conservative: 0  
Best Local Similarity: 99.49% Mismatches: 1  
Query Match: 99.54% Indels: 0  
DB: 9 Gaps: 0  
US-09-966-880A-8 (1-198) x AF529825 (1-596)  
QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
Db 1 ATGGACAGCCTCTTGATGAACCGGAGGAGTTCCTTTACCAATTCAAAAATGTCGCTGG 60  
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
Db 61 GCTAAGGGTCGGGTGAGACCTACCTGTGCTACGTAGTAGAAGAGCGGTGACAGTGTCTACA 120

```
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCCTTTTACAGCTTTGGTTATCTTCGCAATAAGACGGCTGCCACCTGGAATTGCTC 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCCCTCCGCTACATTCGGACTGGGACCTAGACCTGGCGCTGCTACCGCTCACCTGG 240
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCGAGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTCTGTGAGGACCGCAAG 360
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCGAGGGCTGCGCGGCTGCACCGCGGGGTGCAATAGCCATCATGACC 420
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAAGATTATTTTACTGCTGGAAATACTTTTGTAGAAAACCATGAAGAATTTTCAA 480
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGCATCTTT 540
QY 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 541 TTGCCCTCTGATGAGGTTGATGACTTACGAGACGCAATTTCTGACTTTGGGACTT 594

RESULT 15
AF529829 596 bp mRNA linear SYN 19-AUG-2002
LOCUS Mus musculus clone 2 transgenic Homo sapiens AID (AID) mRNA,
DEFINITION complete cds.
ACCESSION AF529829
VERSION AF529829.1 GI:22297245
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 596)
Martin, A. and Scharff, M.D.
Somatic hypermutation of the AID transgene in B and non-B cells
Unpublished
2 (bases 1 to 596)
Martin, A. and Scharff, M.D.
Direct Submission
Submitted (17-JUL-2002) Cell Biology, Albert Einstein College of
Medicine, 1300 Morris Park Ave. Chanin 404, Bronx, NY 10461, USA
JOURNAL Location/Qualifiers
FEATURES
1..596
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="2"
/cell_type="hybridoma Pl-5"
/transgenic
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..>596
/gene="AID"
1..>596
/gene="AID"
/codon_start=1
/product="AID"
/protein_id="AA095416.1"
/db_xref="GI:22297246"
/translation="MDRLMLNRKFLYDFKNRWAGKRRETYLCYVVKRRDSATSFSL
DFGYLRNKGCHVELLFYIISDWLDLPGRCYRVTFWTSWSPCYDCARHVAFLRGNP"

source
Mus musculus
gene
Mus musculus
CDS
Mus musculus
```

```
BASE COUNT 128 a 163 c 155 g 150 t
ORIGIN
Alignment Scores:
Pred. No.: 1.34e-108 Length: 596
Score: 1081.00 Matches: 197
Percent Similarity: 99.49% Conservative: 0
Best Local Similarity: 99.49% Mismatches: 1
Query Match: 99.54% Indels: 0
DB: 12 Gaps: 0
US-09-966-880A-8 (1-198) x AF529829 (1-596)
QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 1 ATGGAGACAGACTCTTGATGAACCGGAGAAAGTTCTTTTACCAATTCAAAAATGTCGGTGG 60
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 61 GCTAAGGCTCGGCTGAGACCTACCTGTGCTAGTAGTAGAAGAGCGGTGACAGTGTCTACA 120
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 121 TCCTTTTCACTGACGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 180
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 181 TTCCCTCCGCTACATTCGGACCTGGACCTAGACCTTGGCGCTGCTACCGCTCACCTGG 240
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 241 TTCACCTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCTGCGAGGG 300
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db 301 AACCCCAACCTCAGTCTGAGGATCTTACCGCGCGCTCTACTCTGTGAGGACCGCAAG 360
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 361 GCTGAGCCGAGGGCTGCGCGGCTGCACCGCGGGGTGCAATAGCCATCATGACC 420
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 421 TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAACCATGAAGAATTTTCAA 480
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 481 GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGCATCTTT 540
QY 181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 541 TTGCCCTCTGATGAGGTTGATGACTTACGAGACGCAATTTCTGACTTTGGGACTT 594
```

Search completed: June 19, 2003, 18:39:41  
Job time : 2488 secs





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 19, 2003, 11:52:06 ; Search time 227 Seconds  
(without alignments)  
1964.298 Million cell updates/sec

Title: US-09-966-880A-8

Perfect score: 1086

Sequence: 1 MDSLMMRRKKFLYQFNVRW.....ILLPLYVDLDRFRTGL 198

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US0966880/runat\_14062003\_175645\_16263/app\_query.fasta\_1.391  
-DB=N\_Geneseq\_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -HEAPSIZ=500 -MINLEN=0 -NAXLEN=2000000000  
-USER=US0966880 @CGN\_1.1.263 @runat\_14062003\_175645\_16263 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N\_Geneseq\_101002.\*  
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1086	100.0	2818	21	AAC55312	Mouse activation-1
2	1008	92.8	2440	21	AAC55307	Human activation-1
3	644.5	59.3	6564	21	AAC55314	Human activation-1
4	644.5	59.3	11204	21	AAC55339	Human activation-1
5	505	46.5	271	21	AAC55317	Human activation-1
6	390	35.9	1534	20	AAZ20856	Polynucleotide seq
7	390	35.9	1534	22	AA592933	Human cDNA encodin
8	390	35.9	1534	24	AA909652	Human polynucleoti
9	388	35.7	1446	21	AAA12409	cDNA encoding a hu
10	369.5	34.0	610	19	AAV48231	Human RNA editing
11	369.5	34.0	610	21	AAV72058	cDNA encoding huma
12	369.5	34.0	950	22	AA541420	cDNA encoding nove
13	369.5	34.0	987	21	AAF16264	Human prostate can
14	369.5	34.0	987	22	AA541407	cDNA encoding nove
15	369.5	34.0	1120	22	AA159847	Human polynucleoti
16	369.5	34.0	1143	22	AA158061	Human polynucleoti
17	363.5	33.5	1348	24	ABN96785	Gene #3283 used to
18	349.5	32.2	944	24	ABL99876	Human secretory po
19	336.5	31.0	1055	24	AAD24392	Human RNA metaboli
20	322.5	29.7	716	20	AA215848	Human gene express
21	321.5	29.6	819	22	AA541669	cDNA encoding nove
22	308.5	28.4	2151	24	AA562572	cDNA sequence #359
23	277.5	25.6	1164	24	AAD27214	Human AAD45360 pro
24	274	25.2	675	24	AAD25771	Human APOBEC2 cDNA
25	274	25.2	4812	22	AAK77995	Human immune/haema
26	274	25.2	4812	22	AAK84817	Human immune/haema
27	274	25.2	12600	24	AAD25770	Human APOBEC2 gene
28	271	25.0	892	19	AAV17184	cDNA encoding a no
29	270	24.9	12600	24	AAD25817	Human APOBEC2 gene
30	262	24.1	6131	22	AA542187	Genomic sequence #
31	262	24.1	6131	22	AAK69784	Human immune/haema
32	262	24.1	201143	24	ABK83568	Human DNA differen
33	260	23.9	148	21	AAC55316	Human activation-1
34	259.5	23.9	1567	22	AA193080	Human polynucleoti
35	212	19.5	116	21	AAC55318	Human activation-1
36	209	19.2	2338	22	AAH17654	Human cDNA sequenc
37	203	18.7	454	21	AAC00672	Human secreted pro
38	199	18.3	879	15	AAQ71633	Apo-B RNA editing
39	193	17.8	879	15	AAQ71632	Apo-B RNA editing
40	192	17.7	674	24	AA562024	Porcine muscular s
41	187	17.2	300	20	AA213088	Human gene express
42	182.5	16.8	650	24	ABQ60647	Human colon cancer
43	167	15.4	445	22	AAC91325	Human polynucleoti
44	162.5	15.0	572	22	ABA63759	Human foetal liver
45	162.5	15.0	572	22	ABA30950	Probe #9416 for ge

ALIGNMENTS

RESULT 1

AAC55312

ID AAC55312 standard; CDNA; 2818 BP.

XX AAC55312:

XX 05-FEB-2001 (first entry)

XX Human activation-induced cytidine deaminase encoding cDNA SEQ ID NO:7.

XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW Immune related disease; allergy; allergic disease; anti-allergic;  
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatologic;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; Pigeege disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;



DR WPI; 2000-611715/58.  
DR P-PSDB; AAB24197.  
XX  
PT Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
XX  
PS Claim 3; Page 126-130; 174pp; Japanese.  
XX  
XX The present sequence encodes mouse activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
CC antianemic, antiasthmatic, ophthalmological, anti-HIV and  
CC dermatological activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class) class  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders.  
XX  
SQ Sequence 2440 BP; 706 A; 546 C; 551 G; 636 T; 1 other;

Alignment Scores:  
Pred. No.: 7,86e-113 Length: 2440  
Score: 1008.00 Matches: 183  
Percent Similarity: 95.94% Conservative: 6  
Best Local Similarity: 92.89% Mismatches: 8  
Query Match: 92.83% Indels: 0  
DB: 21 Gaps: 0

US-09-966-880A-8 (1-198) x AAC55307 (1-2440)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTTP 20  
DB 93 ATGGACAGCCTTCTGATGAGCAAAAGAGTTTCTTACCATTTCAAAATGTCGCTGG 152

QY 21 AlaLysGlyArgArgGlyThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
DB 153 GCCAAGGAGCGCATGAGACCTACCTCTGCTACGTGGTGAAGAGGAGAGATAGTGCCACC 212

QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
DB 213 TCCTGCTCACTGACTTCGGCCACCTTCGCAACAAGTCTGGCTGCCACGTGGAATGTTG 272

QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
DB 273 TTCCTACGCTACATCTCAGACTGGACCTGGACCGCGCGGTGTACCGCGCTCACCTGG 332

QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
DB 333 TTCACCTCCTGGAGCCGCTGCTATGACTGTGCCCGCAGCTGGCTGAGTTCTGAGATGG 392

QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
DB 393 AACCTAACCTCAGCCTGAGGATTTTCACCGCGCGCTCTACTCTGTGAAGACCGCAAG 452

QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
DB 453 GCTGAGCTGAGGGCTCGGAGAGACTGCACCGCTGGGGTCCAGATCGGGATCATGACC 512

QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
DB 513 TTCAAGAGACTATTTTACTGCTGGAATACATTTGTAGAAAAATCGTGAAGAACTTTCAA 572

QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
DB 573 GCCTGGGAAGGGCTACATGAAATTTCTGTCGGGCTAACAGACAACCTTCGGGGCATCTT 632

QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGly 197  
DB 633 TTGCCCTTTGACGAAGTCGATGACTTGGGAGATGCATTTCGTATGTGTGGGA 683

RESULT 3  
AAC55314  
ID AAC55314 standard; DNA; 6564 BP.  
XX  
AC AAC55314;  
XX  
DT 05-FEB-2001 (first entry)  
XX  
DE Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:10.  
XX  
XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; antiallergic;  
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200058480-A1.  
XX  
PD 05-OCT-2000.  
XX  
PF 28-MAR-2000; 2000WO-JP01918.  
XX  
PR 29-MAR-1999; 99JP-0087192.  
PR 24-JUN-1999; 99JP-0178999.  
PR 27-DEC-1999; 99JP-0371382.  
XX  
XX (NISR ) JAPAN TOBACCO INC.  
XX (HONJ/) HONJO T.  
PI HonJo T, Muramatsu M;  
XX  
XX WPI; 2000-611715/58.  
XX  
XX Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -  
XX  
PS Claim 17; Page 145-150; 174pp; Japanese.

XX The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
CC antianemic, antiasthmatic, ophthalmological, anti-HIV and  
CC dermatological activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class) class  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders. The  
XX present sequence represents a genomic DNA sequence of human AID.  
SQ Sequence 6564 BP; 1909 A; 1358 C; 1383 G; 1914 T; 0 other;

Alignment Scores:  
Pred. No.: 1.01e-67 Length: 6564

Score:	644.50	Matches:	177
Percent Similarity:	24.02%	Conservative:	0
Best Local Similarity:	24.02%	Mismatches:	2
Query Match:	59.35%	Indels:	560
DB:	21	Gaps:	2

US-09-966-880A-8 (1-198) x AAC55314 (1-6564)

QY	3	SerLeuLeuMetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTirAlaLys	22
Db	1063	AGCCTCTTGATGAACCGAGGAAGTTCTTTACCAATTCAAAAATGTCCGCTGGGCTAAG	1122
QY	23	GlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPhe	42
Db	1123	GGTCGGCGTGAACCACTACCTGTGCTACGTAGTGAAGAGCGGTGACAGTGTACATCCTTT	1182
QY	43	SerLeuAspPheGlyTyrLeuArgAsnLys	52
Db	1183	TCACCTGGACTTTGGTTATCTTCGCANATAA-CGTATCAATTAAAGTCAAGTCTGCAAGCAG	1241
QY	52	-----	52
Db	1242	TTTAAATGTCAACTGTGAGTGCCTTTTAGAGCCACCTGCTGATGGTATTACTTCCATCCTT	1301
QY	52	-----	52
Db	1302	TTTTGGCATTTGTGCTCTATCATACATTCCTCAAAATCCTTTTTTATTCTTTTCCCATG	1361
QY	52	-----	52
Db	1362	TCCATGCACCATATTAGACATGGCCAAAATATGTGATTAAATTCCTCCCCAGTAATGC	1421
QY	52	-----	52
Db	1422	TGGGACCCCTAATACCACCTCTCCCTTCAGTGCCAGAGAACACTGCTCCCAAACGTTTA	1481
QY	52	-----	52
Db	1482	CCAGCTTTCCCTCAGCATCTGNAATGCGCTTGCCAGCAAAAAATTTAAATGTGAAAAACAAATTCGTGC	1541
QY	52	-----	52
Db	1542	TGGAGAATATATTACGCTTGTCCAGCAAAAAATTTAAATGTGAAAAACAAATTCGTGC	1601
QY	52	-----	52
Db	1602	TTAAGCATTTTGNAAATTAAAGGAAGAAGAAATTTGGGAAAAAATTAAACGGTGGTCAATT	1661
QY	52	-----	52
Db	1662	CTGTTTTCCAAATGATTTCTTTTCCCTCTACTCACATGGGTGCTAGGCCAGTGAATACA	1721
QY	52	-----	52
Db	1722	TTCAACATGTGTATCCCAAGAAACTCAGAGAAGCGCTCGGCTGATGATTAAATTAATTA	1781
QY	52	-----	52
Db	1782	TCITTGGCTACCCGAGAGAAATTACATTTTCCAAGAGACTTCTTCACCAAAAATCCAGATGG	1841
QY	52	-----	52
Db	1842	GTTTACATAAACTTCTGCCCATGGGTATCTCTCTCTCTACACGCTGTGACGCTCGGG	1901
QY	52	-----	52
Db	1902	CTTTGGTGAATCTCAGGGAAGCATCCGTGGGTGGAAGGTCATCGTCTGGCTCGTTGTTT	1961
QY	52	-----	52
Db	1962	GATGGTTATATTACCATGCAATTTTCTTTGGCTACATTTGTATTGAATACATCCCAATCT	2021
QY	52	-----	52

Db 3101 AACTGCTCTTCAACCCCATATACCCGCGCTTTTTTTCCTTTTTTTTTTTTTTTTGAAGATTAT 3160  
Qy 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164  
Db 3161 TTTTACTGCTGGAATACITTTGTAGAAACACGAAAGAACTTTCAAGCCCTGGGAGGG 3220  
Qy 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeu 181  
Db 3221 CTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTGGCGCATCTTTTG 3271

RESULT 4

AAC55339  
ID AAC55339 standard; DNA; 11204 BP.  
XX AAC55339;  
XX  
DT 05-FEB-2001 (first entry)  
XX Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:35.  
DE  
XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; anti-allergic;  
KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; DiGeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; Irg subclass selection disorder; ds.  
XX

OS Homo sapiens.

XX WO200058480-A1.

PN 05-OCT-2000.

PD 28-MAR-2000; 2000WO-JP01918.

XX 29-MAR-1999; 99JP-0087192.

PR 24-JUN-1999; 99JP-0178999.

PR 27-DEC-1999; 99JP-0371382.

XX (NIBS) JAPAN TOBACCO INC.

PA (HONJ/) HONJO T.

XX Honjo T, Muramatsu M;

XX WPI; 2000-611715/58.

XX Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
PT allergies -

XX Claim 17; Page 163-170; 174pp; Japanese.

XX The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
CC has cytidine activity similar to APOBEC-1. AID has anti-allergic,  
CC dermatological activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders. Immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class) class  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC IgE disorder, and Irg subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders. The  
CC present sequence represents a genomic DNA sequence of human AID.

XX SQ Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 other;  
Alignment Scores:  
Pred. No.: 2.18e-67 Length: 11204  
Score: 644.50 Matches: 177  
Percent Similarity: 24.02% Conservative: 0  
Best Local Similarity: 24.02% Mismatches: 2  
Query Match: 59.35% Indels: 560  
DB: 21 Gaps: 2  
US-09-966-880A-8 (1-198) x AAC55339 (1-11204)  
QY 3 SerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLys 22  
Db 6278 AGCCTCTTGATGAACCGGAGGAGTTCITTACCAATTCAAAATGTCCTGGCTAAG 6337  
QY 23 GlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPhe 42  
Db 6338 GGTGGCGGTGAGACCTACCTGTCTAGTAGTGAAGAGCGGTGACAGTGCTACATCCTT 6397  
QY 43 SerLeuAspPheGlyTyrLeuArgAsnLys----- 52  
Db 6398 TCACCTGGACTTGGTTATCTTCGCAATAA--GGTATCAATPAAAGTCAGCTTTCGAACGAG 6456  
QY 52 ----- 52  
Db 6457 TTTAATGGTCAACTGTGAGTGCTTTTAGAGCCACCTGCTGATGATTACTTCCATCCTT 6516  
QY 52 ----- 52  
Db 6517 TTTTGGCATTTGTGCTCTATCATCATTCCTCAAAATCCTTTTCTTTTATTTTCTTCCATG 6576  
QY 52 ----- 52  
Db 6577 TCCATGCCCATATTAGACATGGCCCAAAATATGTGATTAAATTCCTCCCGCAATAATGC 6636  
QY 52 ----- 52  
Db 6637 TGGGCACCCCTAATACCACCTCTTCTTCAGTGCCAGAACAACACTGCTCCCAAACTGTTA 6696  
QY 52 ----- 52  
Db 6697 CCAGCTTTCCTCAGCATCTGAATTGCTTTGAGATTAATTAAGCTAAAGCATTTTATA 6756  
QY 52 ----- 52  
Db 6757 TGGGAGAATATTATCAGCTTGTCCAAGCAAAATTTTAAATGTGAAAAACAATAATTGTGTC 6816  
QY 52 ----- 52  
Db 6817 TTAAGCATTTTGAATAATTAAGGAAGAGAAATTTGGGAAAAAATTAACGGTGGTTCAATT 6876  
QY 52 ----- 52  
Db 6877 CTGTTTCCAAATGATTTCCTTCCCTCTACTCATCATGGTGTGCTAGGCCAGTGAATACA 6936  
QY 52 ----- 52  
Db 6937 TTCAACATGGTGATCCCGAGAAAACTCAGAGAAAGCCTCGGCTGATGATTAAATAATTGA 6996  
QY 52 ----- 52  
Db 6997 TCTTTCGGCTACCCGAGAGAAATTAATTTCCAAGAGACTTCTTCCACCAAAATCCAGATGG 7056  
QY 52 ----- 52  
Db 7057 GTTTACATAAACTTCTGCCCATGGGTATCTCTCTCTTAACACAGCTGTGACGCTCTGGG 7116  
QY 52 ----- 52  
Db 7117 CTTGGTGAATCTCAGGAAGCATCCGTGGGTGGAAGGTATCGTCTGGCTCTGTTGTT 7176

Qy 52 ----- 52  
 Db 7177 GATGGTATATTACCATGCAATTTCTTTGCTCTACATTTTGATTGAATACATCCCAATCT 7236  
 Qy 52 ----- 52  
 Db 7237 CTTCTCTATTTCGGTGACATGACACATTTCTATTTCAGAAGGCTTTGATTTTATCAAGCACT 7296  
 Qy 52 ----- 52  
 Db 7297 TTTCAATTTACTTCTCATGGCAGTGCCTATTACTTCTCTTACAATACCCATCTGTCTGCTTT 7356  
 Qy 52 ----- 52  
 Db 7357 ACCAAATCTATTTCCTCTTTTTCAGATCTCTCCCAATGGTCCCTCATATAAAGTGTCTGCT 7416  
 Qy 52 ----- 52  
 Db 7417 CCACCTAGTGTCCAGGTATATTTCACAAATGTTACATCAACAGGCACTTTAGCCATTT 7476  
 Qy 52 ----- 52  
 Db 7477 TCCTTTCTCAAAAGGTGCAAAAGCAACTTTCATAAACACAAATTAATCTTCGGTGAGGTA 7536  
 Qy 52 ----- 52  
 Db 7537 GTGTGATGCTTCTCTCCCACTCAGCGCACTTCTCTTCTCTCAATTCACAAAAACCCA 7596  
 Qy 52 ----- 52  
 Db 7597 TAGCCTCTCTTCACTCTGCAGGACTAGTGTCTGCCAAGGTTTCAGCTCTACCTACTGGTGT 7656  
 Qy 52 ----- 52  
 Db 7657 GCTCTTTTGAGCAAGTTGCTTAGCCTCTCTGTAAACACAGGACAATAGTGTCAAGCATCC 7716  
 Qy 52 ----- 52  
 Db 7717 CCAAAGATCATTCGAGGAGACAATGACTAAGGCTACCGAGCGCAATAAAAGTCAGTGA 7776  
 Qy 53 ----- AsnGlyCysHisValGluLeuPheLeu 62  
 Db 7777 ATTTTAGCGTGGTCTCTCTCTCTCTCCAGAACGGCTGCACGTGGAAATGCTCTTCCCTC 7836  
 Qy 63 ArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThr 82  
 Db 7837 CGCTACATCTCGACTGGGACCTAGACCTGCGCGTGTACCGGTCACTGGTTCACC 7896  
 Qy 83 SerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnPro 102  
 Db 7897 TCCTGGAGCCCTGCTTACGACTGTGCCGACATGTGGCGACTTTCTGCGAGGGAACCC 7956  
 Qy 103 AsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGlu 122  
 Db 7957 AACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTGTGAGGACCGCAAGGCTGAG 8016  
 Qy 123 ProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLys 142  
 Db 8017 CCGGAGGGCTGCGGGGGTGTGACCGCGCGGGGTGCAATAGCCATCATGACCTTCAA- 8075  
 Qy 142 ----- 142  
 Db 8076 AGGTGCGAAAGGGCTTTCCGCGCAGCGCGAGTGCAGACGCCGCAATTCGGGATTGGGATG 8135  
 Qy 142 ----- 142  
 Db 8136 CGGAATGAATGAGTTAGTGGGAAGCTCGAGGGGAAGAAGTGGGGGGGATTTCTGGTTCA 8195  
 Qy 142 ----- 142  
 Db 8196 CCTCTGGACCCGAAATTAAGATTAGAACGAGAAAAAGAGTGAATGGCTCAGAGCAAG 8255  
 Qy 142 ----- 142

Db 8256 GCCCCGAGGAATGAGAAAATGGGCCAGGTTGCTTCTTCCCTCGATTGGAACCTG 8315  
 Qy 143 ----- AspTyr 144  
 Db 8316 AACTGTCTTCTACCCCATATCCCCGCTTTTTCCTCTTTTTCCTTTTTCCTTTTTCCTTTT 8375  
 Qy 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164  
 Db 8376 TTTTACTGCTGGAATACTTTTGTAGAAAACCCAGAAAGAACTTTTCAAGCCTGGGAAGG 8435  
 Qy 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeu 181  
 Db 8436 CTGTCGATAAATTCAGTTCGTCTCTCCACAGACTTCGCGGCATCCTTTTG 8486  
 RESULT 5  
 AAC55317  
 ID AAC55317 standard; DNA: 271 BP.  
 AC AAC55317;  
 XX  
 DT 05-FEB-2001 (first entry)  
 DE Human activation-induced cytidine deaminase exon 3 SEQ ID NO:13.  
 DE XX  
 KW Activation-induced cytidine deaminase; AID; cytidine deaminase;  
 KW immune related disease; allergy; allergic disease; antiallergic;  
 KW antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;  
 KW gene therapy; B cell associated immune system disorder; food allergy;  
 KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
 KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
 KW drug allergy; allergic rhinitis; rosen disease; Digorge disease; AIDS;  
 KW ataxia telangiectasia; common variable immunodeficiency disorder;  
 KW major histocompatibility class II deficiency disease;  
 KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.  
 XX Homo sapiens.  
 XX WO2000058480-A1.  
 XX  
 PD 05-OCT-2000.  
 XX  
 PF 28-MAR-2000; 2000WO-JP01918.  
 XX  
 PR 29-MAR-1999; 99JP-0087192.  
 PR 24-JUN-1999; 99JP-0178999.  
 PR 27-DEC-1999; 99JP-0371382.  
 XX  
 PA (NISR) JAPAN TOBACCO INC.  
 PA (HONJ/) HONJO T.  
 PI Honjo T, Muramatsu M;  
 XX  
 DR WPI; 2000-611715/58.  
 XX  
 PT Nucleic acid encoding activation induced cytidine deaminase, useful as  
 PT a target for drug development for immune-related diseases including  
 PT allergies -  
 XX  
 PS Claim 18; Page 151; 174pp; Japanese.  
 XX  
 CC The present invention describes an activation-induced cytidine deaminase  
 CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
 CC has cytidine activity similar to APOBEC-1. AID has antiallergic,  
 CC antianemic, antiasthmatic, ophthalmological, anti-HIV and  
 CC dermatological activities, and can be used in gene therapy. AID  
 CC polynucleotides are useful in methods for identifying drugs for the  
 CC treatment of B cell associated immune system disorders, immunodeficiency  
 CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
 CC disease, IgA nephritis, gamma-globulinaemia, atopic dermatitis, allergic  
 CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
 CC disease, Digorge disease, ataxia telangiectasia, common variable  
 CC immunodeficiency disorder, MHC (major histocompatibility class)

CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders. The  
CC present sequence represents the exon 3 genomic DNA sequence of human AID.  
XX Sequence 271 BP; 47 A; 95 C; 76 G; 53 T; 0 other;

Alignment Scores:

Pred. No.: 1-23e-52 Length: 271  
Score: 505.00 Matches: 90  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 46.50% Indels: 0  
DB: 21 Gaps: 0

US-09-966-880A-8 (1-198) x AAC55317 (1-271)

QY 53 AsnGlyCysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspPro 72  
Db 1 AACGGCTGGCAGCTGGAATTGCTTCTCCGCTACATCTCGGACTGGGACCTAGACCC 60  
QY 73 GlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArg 92  
Db 61 GGCCGCTGCTACCGCTCACCTGGTTCACCTCTGGAGCCCTGCTACGACTGTGCCGA 120  
QY 93 HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArg 112  
Db 121 CATGTGGCGACTTCTGGGAGGGAACCCCACTCAGTCTGAGGATCTTCACCGGCGC 180  
QY 113 LeuTyrPheCysGluAspArgTysAlaGluProGluGlyLeuArgLeuHisArgAla 132  
Db 181 CTCTACTTCTGTGAGCAGCGCAAGGCTGAGCCGAGGGGCTGCGGCGGTGCACCGGCC 240  
QY 133 GlyValGlnIleAlaIleMetThrPheLys 142  
Db 241 GGGGTGCAATAGCCATCATGACCTTCAAA 270

RESULT 6  
AAZ20856  
ID AAZ20856 standard; cDNA; 1534 BP.  
XX  
AC AAZ20856;

XX  
DT 09-DEC-1999 (first entry)  
XX  
DE Polynucleotide sequence of the lp547\_4 clone.

XX secreted protein; cDNA library; clone; transmembrane protein;  
KW signal sequence cloning; hybridization cloning; gene therapy;  
KW receptor; ds.  
XX  
OS Homo sapiens.

XX  
FH Key Location/Qualifiers  
FT CDS 51..1205  
FT a /\*tag= a  
FT b /\*product= lp547\_4  
FT sig\_peptide 882..980  
FT /\*tag= b  
FT mat\_peptide 981..1205  
FT /\*tag= c

XX WO942470-A1.  
XX  
XX 26-AUG-1999.  
XX 18-FEB-1999; 99WO-0503458.  
XX  
XX 18-FEB-1998; 98US-0075038.  
PR 17-FEB-1999; 99US-0251600.  
XX  
XX (GEMY ) GENETICS INST INC.

XX  
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Merberg D;  
PI Treacy M, Agostino MJ, Steininger RJ;  
XX  
DR WPI: 1999-518580/43.  
DR P-PSDB; AAY42383.

XX New polynucleotides encoding human secreted proteins used for  
therapeutic, diagnostic and research purposes.

XX Claim 16; Page 104; 125pp; English.

XX This is the polynucleotide sequence of the clone lp547\_4, which was  
isolated from a human fetal brain cDNA library using methods which are  
selective for cDNAs encoding secreted proteins, or by identification as  
a secreted or transmembrane protein on the basis of computer analysis of  
the amino acid sequence of the encoded protein.  
CC The PMS and proteins of the invention are predicted to have biological  
activities which would make them suitable for treating, preventing or  
ameliorating medical conditions in humans and animals, although no  
supporting data is given. Suggested activities include nutritional  
activity, cytokine and cell proliferation/differentiation activity,  
immune stimulating (e.g. as vaccines) or suppressing activity,  
hematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic  
and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
activity, cadherin/tumor invasion suppressor activity, and tumor  
inhibition activity. The PMS are also stated to be useful for gene  
therapy. Other activities include inhibiting the growth, infection or  
function of bacteria, fungi, viruses and other parasites; effecting  
CC bodily characteristics such as, e.g. weight, color, skin, etc.,  
CC effecting biorhythms or circadian cycles; enhancing fertility; treatment  
of depression; treatment of pain; hormonal or endocrine activity.

XX Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 other;

Alignment Scores:  
Pred. No.: 1.77e-37 Length: 1534  
Score: 390.00 Matches: 83  
Percent Similarity: 61.62% Conservative: 31  
Best Local Similarity: 44.86% Mismatches: 59  
Query Match: 35.91% Indels: 12  
DB: 20 Gaps: 3

US-09-966-880A-8 (1-198) x AAZ20856 (1-1534)

QY 6 MetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgArg 25  
Db 639 ATGGATCCACCCACATTCACCTTTCAACTTTAAACAATGAACCTTGGTTCAGAGCGCAT 698  
QY 26 GluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSerLeuAsp 45  
Db 699 GAGACTTACCTGTGTTATGAGGTGGAGCGCATGCACATGACACCTGGGTCTGCTGAAC 758  
QY 46 -----PheGlyTyrLeuArgAsnLys-----AsnGlyCys 55  
Db 759 CAGCGCAGGGCTTCTTATGCACACGAGCTCCACATAACACGGTTCTTCTTGAAGCCGC 818  
QY 56 HisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCys 75  
Db 819 CATGCAGAGCTGTGCTTCTCCTGGAGCTGATTCCTTTTGAAGCTGGACCTGGACGAGC 878  
QY 76 TyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAla 95  
Db 879 TACAGGGTTTACCTGCTTCCCTGGAGCCCTTGCCTCAGCTGTGCCCAAGATGCT 938  
QY 96 AspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuPhe 115  
Db 939 AAATTCATTTCAAAAACAAACACGTCGAGCCTGTGCATCTTCACTGCCCGCATCTAT--- 995  
QY 116 CysGluAspArgLysAlaGluProGluGlyLeuArgLeuHisArgAlaGlyValGln 135  
Db 996 ---GATGATCAAGGAAGATGTCAGAGGGGTGCGCACCTTGGCCGAGGCTGGGGCCAAA 1052



Qy 136 IleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHis 155  
 Db 1053 ATTTCAATAATGACATACAGTGAATTAAGCACTGCTGGACACCTTTGTGGACACAG 1112  
 Qy 156 GluArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175  
 Db 1113 GGATGTCCTTCAGCCCTGGATGACTAGATGACGACGACCAAGACCTGAGTGGGAGG 1172  
 Qy 176 LeuArgArgIleLeu 180  
 Db 1173 CTGCGGGCCATTC 1187

RESULT 7

AAS59293

ID AAS59293 standard; cDNA; 1534 BP.

XX AAS59293;

XX 16-JAN-2002 (first entry)

XX Human cDNA encoding a secreted protein lp547\_4.

XX Human; secreted protein; ss; antiinflammatory; immunosuppressive;

XX cytostatic; neuroprotective; antiarthritic; antimicrobial; vulnary;

XX vasotropic; antidiabetic; virucide; antifertility; anticonvulsant;

XX antirheumatic; antitumor; antitumor; osteopathic; tranquiliser;

XX immune deficiency; severe combined immunodeficiency; SCID; tumour;

XX graft-versus-host disease; multiple sclerosis; cell proliferation;

XX periodontal disease; osteoporosis; osteoarthritis; Alzheimer's disease;

XX Parkinson's disease; Huntington's disease; infection; cardiac disease;

XX stroke; sepsis; inflammatory bowel disease; contraceptive; immunogen;

XX food supplement.

XX Homo sapiens.

XX WO200175068-A2.

XX 11-OCT-2001.

XX 22-MAR-2001; 2001WO-US09369.

XX 30-MAR-2000; 2000US-0539330.

XX 04-DEC-2000; 2000US-0729674.

XX (GEMV ) GENETICS INST INC.

XX Jacobs K, McCoy JM, Lavallie E, Collins-racie LA, Evans C;

XX Treacy M, Agostino MJ, Steininger RJ, Spaulding V, Wong GG;

XX Clark H, Fechtel K, Merberg D;

XX WPT: 2001-639363/73.

XX P-PSDB; AAU39075.

XX Secreted human proteins, useful as vaccine for treating various

XX diseases such as autoimmune disorders (e.g. multiple sclerosis), and

XX nervous system disorders (e.g. stroke).

XX Disclosure; Page 580; 619pp; English.

XX The invention relates to novel human secreted proteins, the nucleic

XX acids encoding them. The protein may exhibit cytokine, cell proliferation

XX or cell differentiation activity or may induce production of other

XX cytokines in certain cell populations and may exhibit immune stimulating

XX or immune suppressing activity, which is useful for the treatment of

XX various immune deficiencies and disorders e.g. severe combined

XX immunodeficiency (SCID), autoimmune disorders e.g. multiple sclerosis,

XX systemic lupus erythematosus, rheumatoid arthritis, autoimmune pulmonary

XX inflammation. The proteins are also useful in the treatment of diseases

XX and disorders including tissue, skin and organ transplantation and in

CC graft-versus-host diseases (GVHD), in the induction of tumour immunity,

CC myeloid or lymphoid cell deficiencies, wound healing and tissue repair,

CC in the treatment of burns, incisions and ulcers, as well as in treatment

CC of periodontal diseases, osteoporosis or osteoarthritis, mediated by

CC inflammatory processes, diseases of the peripheral nervous system,

CC Alzheimer's, Parkinson's disease, and Huntington's disease,

CC amyotrophic lateral sclerosis, and Shy-Drager syndrome, infections,

CC infarction of cardiac and central nervous system vessel e.g. stroke,

CC sepsis, inflammatory bowel disease, ulcers, bone regeneration. The

CC protein, having activin- or inhibin-related activities is useful as a

CC contraceptive based on the ability of inhibins to decrease fertility in

CC female mammals and decrease spermatogenesis in male mammals. The

CC proteins and nucleic acids are also useful as food supplements. The

CC present sequence encodes a secreted protein of the invention.

XX SQ Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 other;

Alignment Scores:

Pred. No.: 1.77e-37 Length: 1534

Score: 390.00 Matches: 83

Percent Similarity: 61.62% Conservative: 31

Best Local Similarity: 44.86% Mismatches: 59

Query Match: 35.91% Indels: 12

DB: 22 Gaps: 3

US-09-966-880A-8 (1-198) x AAS59293 (1-1534)

Qy 6 MetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgArg 25

Db 639 ATGTATCCACCCACATTCCTTCACTTTAAACATGAACCTTGGGTGACAGGCGCAT 698

Qy 26 GluThrTyrLeuGlyValValLysArgArgAspSerAlaThrSerPheSerLeuAsp 45

Db 699 GAGACTTACCTGTGTTATGAGGTGGAGCGCATGCACAAATGCACACCTGGCTCTGCTGAAC 758

Qy 46 -----PheGlyTyrLeuArgAsnLys-----AsnGlyCys 55

Db 759 CAGCGCAGGGCTTCTATGCAACACAGGCTCCACATTAACACGGTTCTTGAAGGCGCG 818

Qy 56 HisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCys 75

Db 819 CATGCAGAGCTGTCTTCTGGACGTGATTCCTTTTGGAAAGCTGGACCTGGACAGGAC 878

Qy 76 TyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAla 95

Db 879 TACAGGGTTACCTGTCTTACCTCCTGGAGCCCTGCTTCAAGTGTGCCAGGAAATGGCT 938

Qy 96 AspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPhe 115

Db 939 AAATTCATTTCAAAAACAAACACGCTGAGCCTGTGCATCTTCACTGCCCGCATCTAT--- 995

Qy 116 CysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGln 135

Db 996 ---GATGATCAAGGAGATGTGAGGGGGCTGCGCACCTCGCCGAGGCTGGGCCAAA 1052

Qy 136 IleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHis 155

Db 1053 ATTTCAATAATGACATACAGTGAATTAAGCACTGCTGGACACCTTTGTGGACACACAG 1112

Qy 156 GluArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175

Db 1113 GGATGTCCTTCAGCCCTGGATGACTAGATGACGACGACCAAGACCTGAGTGGGAGG 1172

Qy 176 LeuArgArgIleLeu 180

Db 1173 CTGCGGGCCATTC 1187

RESULT 8

ABA90962

ID ABA90962 standard; cDNA; 1534 BP.

XX ABA90962;

XX ABA90962;



DT 14-FEB-2002 (first entry)  
 XX Human polynucleotide SEQ ID NO 173.  
 XX  
 KW Human; clone bd306-7; ATCC number 98599; gene therapy;  
 KW immune disorder; bacterial infection; fungal infection; cancer; tumour;  
 KW autoimmune disorder; systemic lupus erythematosus; wound; ulcer; inhibin;  
 KW osteoporosis; osteoarthritis; nervous system disorder; neuropathy;  
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease; activin;  
 KW haemophilia; cardiac infarction; stroke; sepsis; arthritis; vulnerability;  
 KW ischaemia-reperfusion injury; inflammatory bowel disease; chemotactic;  
 KW Crohn's disease; cytostatic; anti-inflammatory; immunomodulatory;  
 KW neuroprotective; haemostatic; thrombolytic; anti-inflammatory; ss.  
 XX Homo sapiens.  
 OS  
 XX US2001039335-A1.  
 PN  
 XX 08-NOV-2001.  
 PD  
 XX  
 XX 04-DEC-2000; 2000US-0729674.  
 XX  
 PR 10-APR-1997; 97US-126425P.  
 PR 04-DEC-1997; 97US-067454P.  
 PR 20-DEC-1997; 97US-068379P.  
 PR 02-JAN-1998; 98US-070346P.  
 PR 07-JAN-1998; 98US-070643P.  
 PR 08-JAN-1998; 98US-070755P.  
 PR 13-JAN-1998; 98US-071304P.  
 PR 22-JAN-1998; 98US-072134P.  
 PR 30-JAN-1998; 98US-073095P.  
 PR 18-FEB-1998; 98US-075038P.  
 PR 30-MAR-2000; 2000US-0539330.  
 PR 23-NOV-1998; 98US-0197886.  
 XX  
 PA (JACO/) JACOBS K.  
 PA (MCCO/) MCCOY J M.  
 PA (LAVA/) LAVALLIE E R.  
 PA (COLL/) COLLINS-RACIE L A.  
 PA (EVAN/) EVANS C.  
 PA (MERB/) MERBERG D.  
 PA (TREB/) TREACY M.  
 PA (AGOS/) AGOSTINO M J.  
 PA (STEI/) STEININGER R J.  
 PA (SPAU/) SPAULDING V.  
 PA (WONG/) WONG G G.  
 PA (CLAR/) CLARK H.  
 PA (FECH/) FECHTEL K.  
 XX  
 PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;  
 PI Wong GG, Clark H, Fechtel K;  
 XX  
 XX WPI: 2002-040725/05.  
 DR P-PSDB: ABB55784.  
 XX  
 XX New secreted proteins and encoding polynucleotides, useful in gene  
 PT therapies, particularly for preventing or treating autoimmune  
 PT disorders, cancer, graft-versus-host disease, wound, osteoporosis,  
 PT stroke or inflammations  
 XX  
 PS Disclosure: Page 310-311; 349pp; English.  
 XX  
 XX The invention relates to isolated polynucleotides (ABA90876-NBA90968 and  
 CC ABA90980) and encoded proteins (ABB55698-ABB55800), especially  
 CC polynucleotides SEQ ID NO 1 (ABA90876) and SEQ ID NO 19 (ABA90885) and  
 CC proteins SEQ ID NO 2 (ABB55698) and SEQ ID NO 20 (ABB55707) contained in  
 CC clones bd306-7 and yb8-1 respectively and the clones bd306-7 and yb8-1  
 CC are deposited with the American Type Culture Collection (ATCC) with  
 CC accession number 98599. The polynucleotides and encoded polypeptides have  
 CC cytostatic, anti-inflammatory, immunomodulator, vulnerability,  
 CC neuroprotective, activin, inhibin, chemotactic, haemostatic, thrombolytic  
 CC and anti-inflammatory activity and acting as cytokine modulators,

CC haematopoiesis regulators, tissue growth modulators and/or cadherin  
 CC suppressors. The polypeptides and polynucleotides are useful in gene  
 CC therapies, particularly for preventing, treating or ameliorating any of  
 CC the following diseases: immune deficiency and disorders; e.g. bacterial  
 CC or fungal infections, autoimmune disorders, cancer, systemic lupus  
 CC erythematosus or graft-versus-host disease; myeloid or lymphoid cell  
 CC deficiencies; wound, burns, incisions and ulcers, osteoporosis or  
 CC osteoarthritis; central and peripheral nervous system diseases and  
 CC neuropathies, e.g. Alzheimer's, Parkinson's disease, Huntington's  
 CC disease, amyotrophic lateral sclerosis or Shy-Drager syndrome;  
 CC haemophilia, cardiac infarction or stroke; inflammations, shock, sepsis  
 CC or systemic inflammatory response syndrome, ischaemia-reperfusion  
 CC injury, endotoxin lethality, arthritis, inflammatory bowel disease or  
 CC Crohn's disease; or tumours or cancers, pemphigus vulgaris or pemphigus  
 CC folliculosis.  
 XX  
 XX Sequence 1534 BP; 495 A; 363 C; 345 G; 331 T; 0 other;  
 SQ  
 Alignment Scores:  
 Pred. No.: 1.77e-37 Length: 1534  
 Score: 390.00 Matches: 83  
 Percent Similarity: 61.62% Conservative: 31  
 Best Local Similarity: 44.86% Mismatches: 59  
 Query Match: 35.91% Indels: 12  
 DB: 24 Gaps: 3  
 US-09-966-880A-8 (1-198) x ABA90962 (1-1534)  
 QY 6 MetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgArg 25  
 DB 639 ATGGATCCACCCACATTCACCTTTCAACATGTAACATGCTTGGGTTCAGAGGACGGCAT 698  
 QY 26 GluThrTyrLeuCystTyrValValLysArgArgAspSerAlaThrSerPheSerLeuAsp 45  
 DB 699 GAGACTTACCTGTGTATGAGGTGGAGCGCATGCACATGACACCTCGGTGCTCGTGAAC 758  
 QY 46 -----PheGlyTyrLeuArgAsnLys-----AsnGlyCys 55  
 DB 759 CAGCGCAGGGCTTTCTCTATGCACACGCTCCACATAACACGCTTTCTTGAAGGCCGC 818  
 QY 56 HisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCys 75  
 DB 819 CATGCAGAGCTGTCTCTCTGGACGTGATTCCTTTTGGAGAGCTGGACCTGGACCGAC 878  
 QY 76 TyrArgValThrTrpPheThrSerTrpSerProCystTyrAspCysAlaArgHisValAla 95  
 DB 879 TACAGGGTTACCTGTCTTACCTCCTCGAGCCCTGCTTACGTGTGCCCGAGGAATGGCT 938  
 QY 96 AspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPhe 115  
 DB 939 AAATTCATTTCAAAAACAAACACAGCTGAGCCTGTGCATCTTCACTGCCCGCATCTAT--- 995  
 QY 116 CysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGln 135  
 DB 996 ---GATGATCAAGGAAGATGTCTAGAGGGGCTCGCACCCCTGCGCCGAGGCTGGGCCAAA 1052  
 QY 136 IleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHis 155  
 DB 1053 ATTTCATTAATGACATACAGTGAATTAAGCACTGCTGGGACCTTTGTGGGACCCACAG 1112  
 QY 156 GluArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175  
 DB 1113 GGATGTCCTTCCAGCCCTGGGATGACTAGATGACACGCCACCGACCTGAGTGGGAGG 1172  
 QY 176 LeuArgArgIleLeu 180  
 DB 1173 CTGCGGGCCATCTC 1187  
 RESULT 9  
 ABA12409  
 ID AAA12409 standard; cDNA; 1446 BP.  
 XX  
 AC AAA12409;

XX 25-JUL-2000 (first entry)  
 DT cDNA encoding a human RNA-associated protein.  
 DE  
 DE Human; RNA-associated protein; cell proliferation; cancer; inflammation;  
 KW immune response; reproductive disorder; actinic keratosis;  
 KW atherosclerosis; arteriosclerosis; bursitis; cirrhosis; hepatitis;  
 KW mixed connective tissue disease; myelofibrosis; primary thrombocythemia;  
 KW paroxysmal nocturnal hemoglobinuria; polycythemia vera; psoriasis;  
 KW trauma; ss.  
 XX  
 OS Homo sapiens.  
 FH  
 FH Key Location/Qualifiers  
 FT 282..1446  
 FT CDS /\*tag= a  
 FT /product= "RNA-associated protein"  
 FT  
 XX W0200015799-A2.  
 XX  
 XX 23-MAR-2000.  
 XX  
 XX 17-SEP-1999; 99NO-US21688.  
 XX  
 XX 17-SEP-1998; 98US-0156039.  
 XX 22-SEP-1998; 98US-0158720.  
 PR 04-NOV-1998; 98US-0186815.  
 PR 08-APR-1999; 99US-0128660.  
 XX  
 XX (INCY-) INCYTE PHARM INC.  
 XX  
 XX Tang YT, Corley NC, Guegler KJ, Gorgone GA, Patterson C;  
 PI Hillman JL, Baughn MR, Lal P, Azimzai Y, Yue H, Yang J;  
 XX  
 DR WPI: 2000-271437/23.  
 DR P-PSDB; AAY84437.  
 XX  
 XX New polypeptides and polynucleotides, useful for preventing and  
 PT treating a disorder associated with increased or decreased expression  
 PT of RNA associated proteins -  
 XX  
 XX Claim 9; Page 119; 131pp; English.  
 XX  
 XX The present sequence encodes a human RNA-associated protein. The  
 CC expression of RNA-associated proteins is closely associated with  
 CC reproductive tissues, nervous tissues, cell proliferation including  
 CC cancer, inflammation and immune responses, and so they may be used  
 CC for diagnosis, treatment or prevention of cell proliferative,  
 CC immune/inflammatory disorders, and reproductive disorders. Diseases  
 CC and disorders which may be treated include actinic keratosis,  
 CC atherosclerosis, arteriosclerosis, bursitis, cirrhosis, hepatitis,  
 CC mixed connective tissue disease, myelofibrosis, paroxysmal nocturnal  
 CC hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia  
 CC and cancers, and trauma.  
 XX  
 SQ Sequence 1446 BP; 370 A; 374 C; 393 G; 308 T; 1 other;  
 Alignment Scores:  
 Pred. No.: 2,86e-37 Length: 1446  
 Score: 388.00 Matches: 82  
 Percent Similarity: 61.62% Conservative: 32  
 Best Local Similarity: 44.32% Mismatches: 59  
 Query Match: 35.73% Indels: 12  
 DB: 21 Gaps: 3  
 US-09-966-880A-8 (1-198) x AAI12409 (1-1446)  
 QY 6 MetAsnArgLysPheLeuTyrGlnPheLysAsnValArgTrrAlaLysGlyArgArg 25  
 DB 870 ATGGATCCACCAACATTCACCTTCAACATTAACATGAACCTTGGGTACAGAGCGGCAT 929  
 QY 26 GluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSerLeuAsp 45

Db 930 GAGACTTACCTGTTATGAGGTGGAGCGCATGCACAATGACACCTGGGTCTGCTGTAAC 989  
 QY 46 -----PheGlyTyrLeuArgAsnLys-----AsnGlyCys 55  
 Db 990 CAGCGCAGGGCTTTCTATGCAACCCAGGCTCCACATAAACACGGTTTCCTTGAAGGCCGC 1049  
 QY 56 HisValGluLeuLeuPheLeuArgTyrIleSerAspTrrAspLeuAspProGlyArgCys 75  
 Db 1050 CATGCAGAGCTGTGCTTCTCTGGACGTGATTCCTTTTGGAAAGCTGGACCTGGACCAGGAC 1109  
 QY 76 TyrArgValThrTrrPheThrSerTrrPserProCysTyrAspCysAlaArgHisValAla 95  
 Db 1110 TACAGGGTTACCTGCTTACCTCCTGGAGCCCTGCTTCACTGCTGCCCGAAATGGCT 1169  
 QY 96 AspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPhe 115  
 Db 1170 AAATTCATTTCAAAAAACAAACACGTCGACCTTTCATCTTCACTGCCCGCATCATAT--- 1226  
 QY 116 CysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGln 135  
 Db 1227 ---GATGATCAAGGAAGATGTCAGGAGGGCTGGCCACCTGGCCGAGCTGGGCCCAA 1283  
 QY 136 IleAlaIleMetThrPheLysAspTyrPheTyrCysTrrAsnThrPheValGluAsnHis 155  
 Db 1284 ATTTCAATACTGACATACAGTGAATTTAAGCAGCTGCTGGACACCTTTTGTGGACCACAG 1343  
 QY 156 GluArgThrPheLysAlaTrrPgluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175  
 Db 1344 GGATGTCCCTCCAGCCCTGGGATGGACTAGAGGAGACACAGCCCAAGCCCTGAGTGGGAGG 1403  
 QY 176 LeuArgArgIleLeu 180  
 Db 1404 CTGCGGGGCATTCTG 1418  
 RESULT 10  
 AAV48231  
 ID AAV48231 standard; cDNA; 610 BP.  
 XX  
 AC AAV48231;  
 XX  
 DT 09-NOV-1998 (first entry)  
 XX  
 DE Human RNA editing enzyme nucleotide sequence.  
 KW ss; human; RNA editing enzyme; REE; pharmaceutical carrier; cancer;  
 KW viral disease; circulatory system disorder; RNA processing;  
 KW hypercholesterolaemia; alpha-galactosidase; apolipoprotein B.  
 OS Homo sapiens.  
 FH  
 FH Key Location/Qualifiers  
 FT 1..573  
 FT CDS /\*tag= a  
 FT /product= "RNA editing enzyme"  
 FT  
 XX US5804185-A.  
 PN  
 XX 08-SEP-1998.  
 PD  
 XX 13-MAR-1997; 97US-0816241.  
 PF  
 XX 13-MAR-1997; 97US-0816241.  
 PR  
 XX (INCY-) INCYTE PHARM INC.  
 PA  
 XX Bandman O, Goli SK;  
 PI  
 XX WPI: 1998-505585/43.  
 DR P-PSDB; AAW77092.  
 XX  
 PT Human RNA editing enzyme and polynucleotide(s) encoding it - useful  
 for recombinant production of the enzyme and treatment and detection

PT of disorders associated with incorrect RNA processing

XX Claim 4; Fig 1; 27pp; English.

XX The human RNA editing enzyme (REE) is used in a pharmaceutical carrier  
XX for the treatment of cancer, viral diseases and circulatory system  
XX disorders. The enzyme is used in vivo for the correct processing of RNA  
XX transcripts of genes e.g. change of a codon in apolipoprotein B (apoB)  
XX RNA to give a 100 and 48 kDa product transcribed from the same  
XX gene. Certain disorders have been linked to incorrect RNA editing, e.g.  
XX failure of apo B editing leads to excessive apoB 100 production and  
XX hypercholesterolaemia. Other disorders thought to be linked to incorrect  
XX RNA processing include aberrant alpha-galactosidase processing in  
XX Fabry's disease and neurofibromatosis type I. The enzyme can be produced  
XX recombinantly to treat related disorders. It can also be used to raise  
XX antibodies for immuno-based detection of REE expression levels  
XX e.g. ELISA.

XX Sequence 610 BP; 158 A; 154 C; 154 G; 144 T; 0 other;

Alignment Scores:

Pred. No.: 1-52e-35 Length: 610  
Score: 369.50 Matches: 79  
Percent Similarity: 58.52% Conservatives: 24  
Best Local Similarity: 44.89% Mismatches: 64  
Query Match: 34.02% Indels: 9  
DB: 19 Gaps: 4

US-09-966-880A-8 (1-198) x AA48231 (1-610)

QY 11 PheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgGluThrTyrLeuCys 30  
Db 49 TTCTACTTCCAAATTTAAAAACCTATGGGAAGCCAGCATCGGAACAACTTGCGTGTGC 108  
QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48  
Db 109 TTCACCGTGAAGGATATAAAGCGCCCTCAGTGTCTCCTGGAAGACG-----GCGTC 162  
QY 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuPheLeuArgTyr 64  
Db 163 TTCGGAACAGGTGATCTCAGACCCATTGTCATGCAGAAAGGTCTCTCTCTTGG 222  
QY 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84  
Db 223 TTCTCGAGGACATCTGTCTCTTAACACAAAGTACCAGGTCACTGTGTACACATCTGG 282  
QY 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104  
Db 283 AGCCCTTGCCAGACTGTGCAGGGAGGTGGCGGAGTTCCTGCCAGGCACAGCAAGTG 342  
QY 105 SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu 124  
Db 343 ATCTCACCATTCTCACCGCCGCTCTACTACTTCCAG--TATCCATGTTACCAAGGAG 399  
QY 125 GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144  
Db 400 GGGCTCCGAGCCTGAGTCAGGAAGGGGTGCTGTGGAGATCATGGACTATGAAGATTTT 459  
QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164  
Db 460 AAATATGTTGGGAAACATTTTGTGTACAAATGATATATGACCCATTCAAGCCTTGGGAAGGA 519  
QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
Db 520 TTAANAACCACTTTCGACTTCTGAAAGAAGCGTACGGAGAGTCTC 567

RESULT 11

AAA72058

ID AAA72058 standard; cDNA; 610 BP.

XX

AC AAA72058;

XX

DT 24-NOV-2000 (first entry)

XX cDNA encoding human RNA editing enzyme REE-2.

XX RNA editing enzyme; REE-2; human; HEPR homologue; REPR homologue;  
XX phorbollin I homologue; cancer; tumour; autoimmune disorder;  
XX circulatory system disorder; hypercholesterolaemia; viral infection;  
XX neurological disease; neurofibromatosis; transcript editing; detection;  
XX ss.

XX Homo sapiens.

XX Key Location/Qualifiers  
XX CDS 1..573  
XX FT /\*tag= a  
XX FT /product= "Human REE-2"

XX US6087108-A.

XX 11-JUL-2000.

XX 03-AUG-1998; 98US-0128395.

XX 13-MAR-1997; 97US-0816241.

XX (INCY-) INCYTE PHARM INC.

XX Goli SK, Bandman O;

XX WPI; 2000-531340/48.

XX P-PSDB; AAB11973.

XX Detecting polynucleotide encoding human RNA editing enzyme comprising  
XX hybridizing an isolated and purified polynucleotide complementary to  
XX the polynucleotide and detecting the hybridization complex -  
XX Claim 5; Fig 1A-B; 27pp; English.

XX This sequence represents the cDNA encoding human RNA editing enzyme  
XX REE-2. cDNA encoding REE-2 was initially isolated in a prostate tumour  
XX cDNA library, with the present sequence representing a consensus. REE-2  
XX has chemical and structural homology with the human apoB mRNA editing  
XX protein HEPR (28% identity), the rat HEPR homologue REPR (30% identity),  
XX and a portion of the mRNA editing enzyme phorbollin I (43% identity).  
XX REE-2 was found to be expressed in a variety of cDNA libraries, a high  
XX proportion of which were derived from tumours, neuronal tissues, immune  
XX system cells or synovial tissue from arthritis patients. REE-2 is  
XX therefore thought to be associated with the development of cancer,  
XX autoimmune disorders, circulatory system disorders (e.g.  
XX hypercholesterolaemia), viral infections and neurological diseases (e.g.,  
XX neurofibromatosis). REE-2 or its nucleic acids may be used in the  
XX diagnosis, treatment and prevention of such diseases via the modulation  
XX of transcript editing, which in turn has effects on the encoded protein  
XX (e.g., an alteration in protein activity). The invention specifically  
XX relates to methods of detecting nucleic acids encoding human REE-2 in a  
XX biological sample.

XX Sequence 610 BP; 158 A; 154 C; 154 G; 144 T; 0 other;

Alignment Scores:

Pred. No.: 1-52e-35 Length: 610  
Score: 369.50 Matches: 79  
Percent Similarity: 58.52% Conservatives: 24  
Best Local Similarity: 44.89% Mismatches: 64  
Query Match: 34.02% Indels: 9  
DB: 21 Gaps: 4

US-09-966-880A-8 (1-198) x AAA72058 (1-610)

QY 11 PheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgGluThrTyrLeuCys 30

Db 49 TTCTACTTCCAAATTTAAAAACCTATGGGAAGCCAGCATCGGAACAACTTGCGTGTGC 108

QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48

```

Db      109  TTACCGTGAAGGTATATAAGCGCCGCTCAGTTGTCTCTGGAAGACG-----GCGCNC 162
Qy      49  LeuArgAsn-----LysAsnGlyCysHisValGluLeuLeuPheLeuArgTyr 64
Db      163  TTCCGAACACAGGTGATCTTGAGACCCATTGTGATGCAGAAAGGTGCTTCCTCTCTGG 222
Qy      65  IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTirpPheThrSerTyr 84
Db      223  TTCTGGGAGGACATATCTCTCTAACAACAAAGTACCACTGCTGACACATCTTGG 282
Qy      85  SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104
Db      283  AGCCCTTGCCAGACTGTCAGGGAGGTGGCGAGTTCCTGCCAGCACACCAACGTG 342
Qy      105  SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu 124
Db      343  AATCTCACCATCTTCAACGCCGCTCTACTACTTCCAG---TATCATGTTACCAGAG 399
Qy      125  GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144
Db      400  GGGCTCCGAGCCTGAGTCAGGAAGGGTCTGTGGAGATCATGGACTATGAAGATTTT 459
Qy      145  PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluCly 164
Db      460  AAATATTGTTGGAAACTTTGTGTACAAATGATATGAGCCATTCAAGCCTTGAAGGGA 519
Qy      165  LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db      520  TTAACCAACTTTCGACTTCTGAAAGAGGCTACGGGAGAGTCTC 567

RESULT 12
AAS41420
ID  AAS41420 standard; cDNA; 950 BP.
XX
AC  AAS41420;
XX
DT  17-DEC-2001 (first entry)
XX
DE  cDNA encoding novel human enzyme polypeptide #636.
KW  Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW  ligase; hyperproliferative disorder; immunodeficiency disorder;
KW  autoimmune disorder; neurological disorder; metabolic disorder;
KW  inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW  blood-related disorder; infectious disorder; gene therapy; cytostatic;
KW  anti arthritic; nephrotropic; anticoagulant; ss.
XX
OS  Homo sapiens.
XX
PN  W0200155301-A2.
XX
PD  02-AUG-2001.
XX
PF  17-JAN-2001; 2001WO-US01239.
XX
PR  31-JAN-2000; 2000US-0179065.
PR  04-FEB-2000; 2000US-0180628.
PR  24-FEB-2000; 2000US-0184664.
PR  02-MAR-2000; 2000US-0186350.
PR  16-MAR-2000; 2000US-0189874.
PR  17-MAR-2000; 2000US-0190076.
PR  18-APR-2000; 2000US-0198123.
PR  19-MAY-2000; 2000US-0205515.
PR  07-JUN-2000; 2000US-0209467.
PR  28-JUN-2000; 2000US-0214886.
PR  30-JUN-2000; 2000US-0215135.
PR  07-JUL-2000; 2000US-0216647.
PR  17-JUL-2000; 2000US-0216880.
PR  11-JUL-2000; 2000US-0217487.
PR  14-JUL-2000; 2000US-0217496.
PR  01-NOV-2000; 2000US-0218290.
PR  26-JUL-2000; 2000US-0220963.
PR  26-JUL-2000; 2000US-0220964.
PR  14-AUG-2000; 2000US-0224518.
PR  14-AUG-2000; 2000US-0224519.
PR  14-AUG-2000; 2000US-0225213.
PR  14-AUG-2000; 2000US-0225214.
PR  14-AUG-2000; 2000US-0225266.
PR  14-AUG-2000; 2000US-0225267.
PR  14-AUG-2000; 2000US-0225268.
PR  14-AUG-2000; 2000US-0225270.
PR  14-AUG-2000; 2000US-0225447.
PR  14-AUG-2000; 2000US-0225757.
PR  14-AUG-2000; 2000US-0225758.
PR  14-AUG-2000; 2000US-0225759.
PR  18-AUG-2000; 2000US-0226279.
PR  22-AUG-2000; 2000US-0226681.
PR  22-AUG-2000; 2000US-0226686.
PR  22-AUG-2000; 2000US-0227182.
PR  23-AUG-2000; 2000US-0227009.
PR  30-AUG-2000; 2000US-0228924.
PR  01-SEP-2000; 2000US-0229287.
PR  01-SEP-2000; 2000US-0229343.
PR  01-SEP-2000; 2000US-0229344.
PR  05-SEP-2000; 2000US-0229345.
PR  05-SEP-2000; 2000US-0229509.
PR  05-SEP-2000; 2000US-0229513.
PR  06-SEP-2000; 2000US-0230437.
PR  06-SEP-2000; 2000US-0230438.
PR  08-SEP-2000; 2000US-0231242.
PR  08-SEP-2000; 2000US-0231243.
PR  08-SEP-2000; 2000US-0231244.
PR  08-SEP-2000; 2000US-0231413.
PR  08-SEP-2000; 2000US-0231414.
PR  08-SEP-2000; 2000US-0232080.
PR  08-SEP-2000; 2000US-0232081.
PR  12-SEP-2000; 2000US-0231968.
PR  14-SEP-2000; 2000US-0232397.
PR  14-SEP-2000; 2000US-0232398.
PR  14-SEP-2000; 2000US-0232399.
PR  14-SEP-2000; 2000US-0232400.
PR  14-SEP-2000; 2000US-0232401.
PR  14-SEP-2000; 2000US-0233063.
PR  14-SEP-2000; 2000US-0233064.
PR  21-SEP-2000; 2000US-0233065.
PR  21-SEP-2000; 2000US-0234223.
PR  25-SEP-2000; 2000US-0234274.
PR  25-SEP-2000; 2000US-0234997.
PR  25-SEP-2000; 2000US-0234998.
PR  26-SEP-2000; 2000US-0235484.
PR  27-SEP-2000; 2000US-0235834.
PR  27-SEP-2000; 2000US-0235836.
PR  29-SEP-2000; 2000US-0236327.
PR  29-SEP-2000; 2000US-0236367.
PR  29-SEP-2000; 2000US-0236368.
PR  29-SEP-2000; 2000US-0236369.
PR  29-SEP-2000; 2000US-0236370.
PR  02-OCT-2000; 2000US-0236802.
PR  02-OCT-2000; 2000US-0237037.
PR  02-OCT-2000; 2000US-0237038.
PR  02-OCT-2000; 2000US-0237039.
PR  02-OCT-2000; 2000US-0237040.
PR  13-OCT-2000; 2000US-0239935.
PR  13-OCT-2000; 2000US-0239937.
PR  20-OCT-2000; 2000US-0240960.
PR  20-OCT-2000; 2000US-0241221.
PR  20-OCT-2000; 2000US-0241785.
PR  20-OCT-2000; 2000US-0241786.
PR  20-OCT-2000; 2000US-0241787.
PR  20-OCT-2000; 2000US-0241808.
PR  20-OCT-2000; 2000US-0241809.
PR  20-OCT-2000; 2000US-0241826.
PR  01-NOV-2000; 2000US-0244617.
PR  08-NOV-2000; 2000US-0246474.
PR  08-NOV-2000; 2000US-0246475.

```

PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 11-DEC-2000; 2000US-0251990.  
PR 05-JAN-2001; 2000US-0254097.  
PR 05-JAN-2001; 2000US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX  
DR  
DR WPI: 2001-465566/50.  
DR P-PSDB; AU233550.  
XX  
PT Novel polypeptides and polynucleotides useful for diagnosing,  
PT preventing, treating neural, immune system, muscular, reproductive,  
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
PT diseases .  
XX  
PS Claim 4; SEQ ID No 646; 1180pp; English.  
XX  
XX  
CC The present invention relates to the isolation of novel human enzyme  
CC polypeptides (AAU22915-AU23814), and the cDNA and genomic sequences  
CC encoding them. The enzyme polypeptides of the invention may comprise the  
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
CC isomerases or ligases. The sequences of the invention are useful in the  
CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
CC disorders including hyperproliferative disorders (e.g. cancer),  
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders  
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
CC blood-related disorders (e.g. haemophilia), reproductive disorders

CC (e.g. infertility) and infectious disorders (e.g. Influenza). The  
CC polynucleotides of the invention can also be used in gene therapy.  
CC AAS40785-AAS41684 represent cDNA sequences encoding for the novel human  
CC enzyme polypeptides of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 950 BP; 219 A; 289 C; 219 G; 220 T; 3 other;  
  
Alignment Scores:  
Pred. No.: 2.88e-35 Length: 950  
Score: 369.50 Matches: 79  
Percent Similarity: 58.52% Conservative: 24  
Best Local Similarity: 44.89% Mismatches: 64  
Query Match: 34.02% Indels: 9  
DB: 22 Gaps: 4  
  
US-09-966-880A-8 (1-198) x AAS41420 (1-950)  
  
QY 11 PheLeuTyrGlnPheLysAsnValArgTrrAlaLysGlyArgGluThrTyrLeuCys 30  
Db 110 TTCTACTTCCAAATTTAAACCTATGGGAAGCCACAGATCGGAACAACTTGGCTGTGC 169  
QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48  
Db 170 TTCACCGTGGAGGTATAAAGCCGCTCAGTTGTCTCTCTGGAGAGCG-----GGCGTC 223  
QY 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuLeuPheLeuArgTyr 64  
Db 224 TTCCGAAACACAGGTGGATCTGAGACCCATTGTGATGCAGAAAGGTCTTCTCTCTCTGG 283  
QY 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84  
Db 284 TTTCGCGACGACATCTGTCTCTAACACAAAGTACCAGGTACCTGGTACATCTTGG 343  
QY 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104  
Db 344 AGCCCTTGCCAGACTGTGCAGGGGAGGTGCCGAGTTCCTGGCCAGCAGCAGCGTG 403  
QY 105 SerLeuArgIlePheThrAlaArgLeuTyrPheCysGlyAspArgLysAlaGluProGlu 124  
Db 404 AATCTCACCATCTTCACCGCCGCTCTACTACTTCCAG--TATCCATGTTTACCAGAG 460  
QY 125 GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144  
Db 461 GGGCTCCGACGCTGAGTCAGGAAGGGGTGCTGTGGAGATCATGGACTATGAAGATTTT 520  
QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164  
Db 521 AAATATTGTTGGGAAAACCTTTGTGTACATGATGATGATGATGATGATGATGATGAT 580  
QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
Db 581 TTAATAACCAACTTTCGACTTCTGAAAGAAAGAGGTACGGGAGAGTCTC 628  
  
RESULT 13  
AAFI6264  
ID AAFI6264 standard; cDNA; 987 BP.  
XX  
XX AAFI6264;  
AC  
XX  
DT 13-MAR-2001 (first entry)  
XX  
XX Human prostate cancer antigen nucleotide sequence SEQ ID NO:699.  
DE  
XX Human prostate cancer; prostate cancer antigen; detection; diagnosis;  
KW neuroprotective; cytostatic; cardioactive; immunomodulatory; muscular;  
KW vulnery; gastrointestinal; nephrotropic; antineoplastic; gynaecological;  
KW antibacterial; gene therapy; immune; reproductive; renal;  
KW gastrointestinal; pulmonary; cardiovascular; proliferative disorder;  
KW wound; infectious disease; ss.  
XX



PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.

PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI  
XX  
DR WPI; 2001-465566/50.  
P-PSDB; AAU23537.  
XX  
PT Novel polypeptides and polynucleotides useful for diagnosing,  
PT preventing, treating neural, immune system, muscular, reproductive,  
PT pulmonary, cardiovascular, renal, proliferative disorders and cancerous  
XX diseases  
PS Claim 4; SEQ ID No 633; 1180pp; English.  
XX  
CC The present invention relates to the isolation of novel human enzyme  
CC polypeptides (AAU22915-AAU23814), and the cDNA and genomic sequences  
CC encoding them. The enzyme polypeptides of the invention may comprise the  
CC functional classes of oxidoreductases, transferases, hydrolases, lyases,  
CC isomerases or ligases. The sequences of the invention are useful in the  
CC diagnosis, treatment, prevention and/or prognosis of a wide range of  
CC disorders, including hyperproliferative disorders (e.g. cancer),  
CC immunodeficiency disorders (e.g. AIDS) autoimmune disorders  
CC (e.g. arthritis), neurological disorders (e.g. Alzheimer's disease),  
CC metabolic disorders (e.g. phenylketonuria), inflammatory disorders  
CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),  
CC blood-related disorders (e.g. haemophilia), reproductive disorders  
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The  
CC polynucleotides of the invention can also be used in gene therapy.  
CC AAU40785-AAU41684 represent cDNA sequences encoding for the novel human  
CC enzyme polypeptides of the invention.  
CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 987 BP; 231 A; 298 C; 232 G; 224 T; 2 other:  
  
Alignment Scores:  
Pred. No.: 3 04e-35 Length: 987  
Score: 369.50 Matches: 79  
Percent Similarity: 58.52% Conservative: 24  
Best Local Similarity: 44.89% Mismatches: 64  
Query Match: 34.02% Indels: 9  
DB: 22 Gaps: 4

US-09-966-880A-8 (1-198) x AAS41407 (1-987)

QY 11 PheLeuTyrGlnPheLysAsnValArgTTrpAlaLysGlyArgArgGluThrTyrLeuCys 30  
 Db 147 TTCTACTTCCAAATTTAAACCTTATGGGAAGCCACGATCGGAACAACTTGGCTGTGC 206  
 QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48  
 Db 207 TTCACCGTGAAGGTATAAAGCGCGCTCAGTTGTCTCTGGAGACG-----GGCGTC 260  
 QY 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuLeuPheLeuArgTyr 64  
 Db 261 TTCGGAACACAGGTGATCTCGAGACCCCATTCATGCAGAAAGGTCTCTCTCTGG 320  
 QY 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84  
 Db 321 TTCTGGAGGACATCTGCTCTTAACACAAAGTACCGAGTCCCTGGTACACATCTGG 380  
 QY 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104  
 Db 381 ACCCTTCCCGAGCTGTCAGGGAGGTGGCGGAGTTCCTGGCCAGGCACGACGTCG 440  
 QY 105 SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu 124  
 Db 441 AATCTCACCATCTTCCAGCGCGCTCTACTACTTCCAG---TATCCATGTTACCGAG 497  
 QY 125 GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144  
 Db 498 GGGTCCCGACGCTGAGTCAGGAAGGGTGGCTGCTGGAGATCATGGATATGAAGATT 557  
 QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164  
 Db 558 AAATATTGTTGGGAAACTTGTGTACAAATGATATGAGCCATCAAGCCTTGAAGGGA 617  
 QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
 Db 618 TTAACAAACCACTTTCGACTTCTGAAAGAAAGGCTACGGGAGAGTCTC 665

RESULT 15

AAI59847/c  
 ID AAI59847 standard; cDNA; 1120 BP.  
 XX AC AAI59847;

XX AC 22-OCT-2001 (first entry)

XX DE Human polynucleotide SEQ ID NO 3836.

XX KW Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia; ss.

XX OS Homo sapiens.

XX PN WO200153312-A1.

XX PD 26-JUL-2001.

XX XX 26-DEC-2000; 2000WO-US34263.

XX PR 21-JAN-2000; 2000US-0489725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX WPI; 2001-442253/47.  
 DR P-PSDB; AAM40691.  
 XX Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX Claim 1; SEQ ID NO 3836; 10078pp; English.

CC The invention relates to human nucleic acids (AAI57798-AAI61369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nontropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: immune system suppression,  
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.

XX SQ Sequence 1120 BP; 250 A; 268 C; 332 G; 270 T; 0 other;

Alignment Scores:

Pred. No.: 3,64e-35 Length: 1120  
 Score: 369.50 Matches: 79  
 Percent Similarity: 58.52% Conservative: 24  
 Best Local Similarity: 44.89% Mismatches: 64  
 Query Match: 34.02% Indels: 9  
 DB: 22 Gaps: 4

US-09-966-880A-8 (1-198) x AAI59847 (1-1120)

QY 11 PheLeuTyrGlnPheLysAsnValArgTTrpAlaLysGlyArgArgGluThrTyrLeuCys 30  
 Db 977 TTCTACTTCCAAATTTAAACCTTATGGGAAGCCACGATCGGAACAACTTGGCTGTGC 918  
 QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48  
 Db 917 TTCACCGTGAAGGTATAAAGCGCGCTCAGTTGTCTCTCTGGAGACG-----GGCGTC 864  
 QY 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuLeuPheLeuArgTyr 64  
 Db 863 TTCGAAACACAGGTGGATTCTTGAGACCCATTGTCTATGCAGAAAGGTGCTCTCTCTTGG 804  
 QY 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84  
 Db 803 TTCTCCGACGACATCTGCTCTTAACACAAAGTACCAGGTCACCTGCTACACATCTGG 744  
 QY 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104  
 Db 743 AGCCCTTGGCCAGACTGTGACGGGAGGTGGCCGAGTTCCTGGCCAGGCACGACGACG 684  
 QY 105 SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu 124  
 Db 683 AATCTCACCATCTTCCAGCGCGCTCTACTACTTCCAG---TATCCATGTTACCGAGG 627  
 QY 125 GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144  
 Db 626 GGGCTCCGACGCTGAGTCAAGGAGGGTGGCTGAGATCATGACTATGAAGATT 567  
 QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164



Db 566 AAATATTGTTGGGAAACTTTGTGTACATGATGATGAGCCATTCAAGCCTTGAAGGGA 507

Qy 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgA<sup>g</sup>IlLeu 180  
||| ||| ||| ::||| |||  
||| ||| ||| ::||| |||

Db 506 TTAARAAACCAACTTTCGACTTCTGAARAGAAGGCTACGGGAGAGTCTC 459

Search completed: June 19, 2003, 17:58:08  
Job time : 233 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 19, 2003, 13:52:31 ; Search time 1384 Seconds  
(without alignments)  
2316.984 Million cell updates/sec

Title: US-09-966-880A-8  
Perfect score: 1086  
Sequence: 1 MDSLMLNRKFLYQFKNRW.....ILLPLYEVDLDAFTLGL 198

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+p2n model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US09966880/runat\_14062003\_175645\_16282/app\_query.fasta\_1.391  
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosom62 -TRANS=human40.cdi -LIST=45  
-DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09966880 @cgn\_1\_1.1906 @runat\_14062003\_175645\_16282 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST: \*  
1: em\_estba: \*  
2: em\_esthum: \*  
3: em\_estin: \*  
4: em\_estmu: \*  
5: em\_estov: \*  
6: em\_estpl: \*  
7: em\_estro: \*  
8: em\_hct: \*  
9: gb\_est1: \*  
10: gb\_est2: \*  
11: gb\_hic: \*  
12: gb\_est3: \*  
13: gb\_est4: \*  
14: gb\_est5: \*  
15: em\_estfun: \*  
16: em\_estom: \*  
17: gb\_gss: \*  
18: em\_gss\_hum: \*  
19: em\_gss\_inv: \*  
20: em\_gss\_pln: \*  
21: em\_gss\_vrt: \*  
22: em\_gss\_fun: \*  
23: em\_gss\_mam: \*  
24: em\_gss\_mus: \*  
25: em\_gss\_other: \*  
26: em\_gss\_pro: \*  
27: em\_gss\_rod: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1086	100.0	743	12	BG686133	BG686133 602638412
2	1086	100.0	856	9	AL559877	AL559877 AL559877
3	1086	100.0	872	12	BG758510	BG758510 602711271
4	1086	100.0	953	14	BQ065440	BQ065440 AGENCOURT
5	1086	100.0	1052	14	BQ055935	BQ055935 AGENCOURT
6	1062	97.8	820	12	BG757089	BG757089 602715124
7	1052	96.9	693	12	BG757392	BG757392 602711022
8	1033	95.1	942	12	BF975166	BF975166 602244657
9	986	90.8	757	9	AJ446140	AJ446140 AJ446140
10	977	90.0	696	9	AJ453647	AJ453647 AJ453647
11	910	83.8	729	9	AJ450296	AJ450296 AJ450296
12	884	81.4	688	9	AJ450317	AJ450317 AJ450317
13	868	79.9	522	12	BG144705	BG144705 ut73f07.y
14	858	79.0	541	12	BF238155	BF238155 601811880
15	850	78.3	650	9	AJ449745	AJ449745 AJ449745
16	780	71.8	889	12	BG686876	BG686876 602650861
17	724	66.7	623	9	AJ450295	AJ450295 AJ450295
18	675	62.2	693	12	BF975096	BF975096 602245679
19	665	61.2	653	10	BG637360	BG637360 BB637360
20	378	34.8	884	13	B1834760	B1834760 603090364
21	371.5	34.2	986	14	BQ877348	BQ877348 AGENCOURT
22	370.5	34.1	1100	14	BM914942	BM914942 AGENCOURT
23	370	34.1	1073	12	BG342101	BG342101 602463048
24	369.5	34.0	741	12	BE886229	BE886229 601509806
25	369.5	34.0	923	12	BG025943	BG025943 602292096
26	369.5	34.0	940	12	BG032405	BG032405 602301392
27	369.5	34.0	974	14	BQ707723	BQ707723 AGENCOURT
28	369.5	34.0	1005	13	BM474553	BM474553 AGENCOURT
29	369.5	34.0	1020	13	BM424095	BM424095 AGENCOURT
30	369.5	34.0	1063	14	BM809617	BM809617 AGENCOURT
31	369.5	34.0	1117	14	BM917194	BM917194 AGENCOURT
32	369.5	34.0	1119	11	BC021080	BC021080 Homo sapi
33	369.5	34.0	1128	13	BM559604	BM559604 AGENCOURT
34	369	34.0	1549	11	BC009683	BC009683 Homo sapi
35	367.5	33.8	782	12	BE882870	BE882870 601509274
36	367.5	33.8	1665	13	BM473129	BM473129 AGENCOURT
37	363.5	33.5	1005	14	BQ056030	BQ056030 AGENCOURT
38	363.5	33.5	1052	14	BQ063928	BQ063928 AGENCOURT
39	363.5	33.5	1083	14	BQ058970	BQ058970 AGENCOURT
40	362.5	33.4	1021	14	BQ061901	BQ061901 AGENCOURT
41	360.5	33.2	772	14	BM935138	BM935138 UI-M-BH3 -
42	357	32.9	1020	14	BQ052514	BQ052514 AGENCOURT
43	356.5	32.8	787	10	BE270707	BE270707 601183701
44	356.5	32.8	1021	14	BQ065295	BQ065295 AGENCOURT
45	356	32.8	836	12	BG744603	BG744603 602722683

ALIGNMENTS

RESULT 1  
BG686133  
LOCUS 602638412F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4766234 5',  
DEFINITION 743 bp mRNA linear EST 01-MAY-2001  
mRNA sequence.  
ACCESSION BG686133  
VERSION BG686133.1 GI:13917530  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 743)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL  
COMMENT

Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-remail.nih.gov  
 Tissue procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLC1626 row: 9 column: 03  
 High quality sequence stop: 740.  
 FEATURES  
 source  
 1. 743  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4766234"  
 /clone\_lib="NIH\_MGC\_48"  
 /tissue\_type="primary B-cells from tonsils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: B-cells; Vector: pOTB7; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dr priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."  
 BASE COUNT 176 a 197 c 188 g 182 t  
 ORIGIN

## Alignment Scores:

Pred. No.: 7,26e-131 Length: 743  
 Score: 1086.00 Matches: 198  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x BG686133 (1-743)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
 DB 78 ATGGACACCCCTCTGTATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCCGCTGG 137  
 QY 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
 DB 138 GCTAAGGGTCGGCTGAGACCTACCTGTCTACGTAGTGAAGCGCGTGACGTGCTACA 197  
 QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
 DB 198 TCCTTTTCACCTGGACTTGGTTATCTTCGCAATAAGAACCGCTGCCACGTGGAATTCGTC 257  
 QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
 DB 258 TTCTCGCTGCTACATCTCGGACTGGACCTAGACCCCTGGCCGCTGCTACCGCGTCACCTGG 317  
 QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
 DB 318 TTCACCTCTGGAGCCCTCTGCTACGACTGTGCCGACATGTGCCGACTTTCGCGAGGG 377  
 QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
 DB 378 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 437  
 QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValcInIleAlaIleMetThr 140  
 DB 438 GCTGAGCCGAGGGGCTGGCGGCTGCAACCGCGGGGTGCAATATGCCATCATGACC 497  
 QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
 DB 498 TTCAAAGATTATTTTACTGCTGGAATATCTTTGTAGAAAACCATGAAAGAACTTTCANA 557

QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgAlaGluLeu 180  
 DB 558 GCCTGGGAAGGGTGCATGAAATTCAGTTCGTCTCTCCASACAGCTTCGGCGCATCCTT 617  
 QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198  
 DB 618 TTGCCCTGTATGAGGTGTATGACTTACGAGACGATTCGTACTTTGGGACTT 671

## RESULT 2

AL559877

LOCUS

DEFINITION

AL559877 LTI\_FL011\_BC1 Homo sapiens cDNA clone CS0DG003YB14 5 prime

ACCESION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Genoscope - Centre National de Sequencage

Bp 191 91006 EVRY cedex - France

Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

Location/Qualifiers

1. 856

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="CS0DG003YB14"

/clone\_lib="LTI\_FL011\_BC1"

/sex="male"

/tissue\_type="B cells from Burkitt lymphoma"

/lab\_host="DH10B"

/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed

with a NotI-oligo(dT) primer. Five prime end enriched,

double-stranded cDNA was digested with Not I and cloned

into the Not I and Eco RV sites of the pCMVSPORT 6 vector.

Library was constructed by Life Technologies. Contact :

Feng Liang Life Technologies, a division of Invitrogen

9800 Medical Center Drive Rockville, Maryland 20850, USA

Fax : (1) 301 610 8371 Email : filiang@life.com

http://fulllength.invitrogen.com"

BASE COUNT 209 a 217 c 202 g 226 t

ORIGIN

Alignment Scores:

Pred. No.: 9,22e-131 Length: 856

Score: 1086.00 Matches: 198

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AL559877 (1-856)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20

DB 19 ATGGACACCCCTCTGTATGAACCGGAGGAGTTCTTTACCAATTCAAAATGTCCGCTGG 78

QY 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40

DB 79 GCTAAGGGTCGGCTGAGACCTACCTGTCTACGTAGTGAAGAGCGGTGACAGTGTACA 138

QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60

DB 139 TCCTTTTCACCTGGACTTGTGTTATCTTCGCAATAAGAACCGCTGCCACGTGGAATTCGCTC 198

QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80

```

199 TTCTCCGCTACATCTCGGACTCGGACCTAGACCTTGCGCGCTGCTACCGCGTCACTGG 258
      |||
81 PheThrSerTrpSerProCysTyrAspCysAlaArqHisValAlaAspPheLeuArgGly 100
      |||
259 TTCACCTCTCGAGCGCCCTGCTACGACTTGCGCCGACATGTGCGGAGCTTCTGCGAGGG 318
      |||
101 AsnProAsnLeuSerLeuArgIlePheThrAlaArqLeuTyrPheCysGluAspArgLys 120
      |||
319 AACCCCAACCTCAGTCTGAGGATCTCACCGCGCGCTCTACTTCTGTGAGGACCCGCAAG 378
      |||
121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
      |||
379 GCTGAGCCGAGGGCTGCGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 438
      |||
141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisLeuArgThrPheLys 160
      |||
439 TTCAAGATATTTTACTGCTGGAAATCTTTGTAGAAACCATGAAAGAACTTTCAAA 498
      |||
161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgIleLeu 180
      |||
499 GCCTGGGAGGCGTCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGGCATCCTT 558
      |||
181 LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
      |||
559 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTTGCTACTTTGGGACTT 612

RESULT 3
BG758510          872 bp  mRNA  linear  EST 15-MAY-2001
LOCUS             602172121F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4853069 5',
DEFINITION        mRNA sequence.
ACCESSION         BG758510
VERSION           BG758510.1 GI:14069163
KEYWORDS          EST.
SOURCE            human.
ORGANISM          Homo sapiens
REFERENCE         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS          Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE            NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL           National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT          Unpublished (1999)
                  Contact: Robert Strausberg, Ph.D.
                  Email: cgabbs-r@mail.nih.gov
                  Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
                  CDNA Library Preparation: Ling Hong/Rubin Laboratory
                  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                  DNA Sequencing by: Incyte Genomics, Inc.
                  Clone distribution: MGC clone distribution information can be
                  found through the I.M.A.G.E. Consortium/LLNL at:
                  http://image.llnl.gov
                  Plate: LLCM1698 row: 1 column: 06
                  High quality sequence stop: 836.
                  Location/Qualifiers
                    1. .872
                      /organism="Homo sapiens"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:4853069"
                      /clone_lib="NIH_MGC_48"
                      /tissue_type="primary B-cells from tonsils (cell line)"
                      /lab_host="PH10B (phage-resistant)"
                      /note="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
                      Site:2: EcoRI; cDNA made by oligo-dT priming.
                      Directionally cloned into EcoRI/XhoI sites using the
                      following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                      for average insert size 1.8kb. Library constructed by Ling
                      Hong in the laboratory of Gerald M. Rubin (University of
                      California, Berkeley) using ZAP-cDNA synthesis kit
                      (Stratagene) and Superscript II RT (Life Technologies).
                      Note: this is a NIH_MGC Library."
211 a 221 c 212 g 228 t

```

```

Alignment Scores: 9.51e-131 Length: 872
Pred. No.: 1086.00 Matches: 198
Score: 100.00% Conservative: 0
Percent Similarity: 100.00% Mismatches: 0
Best Local Similarity: 100.00% Indels: 0
Query Match: 12 Gaps: 0
DB: 12
US-09-966-880A-8 (1-198) x BG758510 (1-872)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
      |||
Db 66 ATGGACAGCGCTCTTGATGAACCGGAGGAAGTTTCTTTACCAATTTCAAAAATGTCGCTGG 125
      |||
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgSerAlaThr 40
      |||
Db 126 GCTAAGGGTGGGGTGGAGACCTACCTGTGCTAGTAGAAGAGCGGTGACAGTGCTACA 185
      |||
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
      |||
Db 186 TCCTTTTCACTGGACCTTGGTTATCTTCGCAATAGAACGGCTGCCACGTGGAATTCGTC 245
      |||
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
      |||
Db 246 TTCTCCGCTACATCTCGGACTCGGACCTAGACCTTGGCCGCTGCTACCGCGTCACCTGG 305
      |||
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
      |||
Db 306 TTCACCTCTCGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACTTTCGCGAGGG 365
      |||
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
      |||
Db 366 AACCCCAACCTCAGTCTGAGGATCTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 425
      |||
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
      |||
Db 426 GCTGAGCCGAGGGCTGCGCGGCTGCACCGCGCGGGTGCAGAAATAGCATCATGACC 485
      |||
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
      |||
Db 486 TTCAAGATATTTTACTGCTGGAATFACITTTGTAGAAACCATGAAAGAACTTTCAAA 545
      |||
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
      |||
Db 546 GCCTGGGAGGCGTGCATGAAATTCAGTTGCTCTCCAGACAGCTTCGCGCATCCTT 605
      |||
QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
      |||
Db 606 TTGCCCCCTGATGAGGTTGATGACTTACGAGACGCAATTTGCTACTTTGGGACTT 659

RESULT 4
BQ065440          953 bp  mRNA  linear  EST 02-APR-2002
LOCUS             AGENCOURT_6855061 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5929977
DEFINITION        5', mRNA sequence.
ACCESSION         BQ065440
VERSION           BQ065440.1 GI:19894486
KEYWORDS          EST.
SOURCE            human.
ORGANISM          Homo sapiens
REFERENCE         Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS          Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
TITLE            NIH-MGC http://mgc.nci.nih.gov/.
JOURNAL           National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT          Unpublished (1999)
                  Contact: Robert Strausberg, Ph.D.
                  Email: cgabbs-r@mail.nih.gov
                  Tissue Procurement: Lou Staudt
                  CDNA Library Preparation: Rubin Laboratory
                  CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                  DNA Sequencing by: Agencourt Bioscience Corporation

```

BASE COUNT  
ORIGIN

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Plate: LLCM2108 row: p column: 10

High quality sequence stop: 634.

#### FEATURES

Location/Qualifiers

1..953  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:592977"  
 /clone\_lib="NIH\_MGC\_99"  
 /tissue\_type="lymphoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lymph; Vector: pOTB7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dt priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH\_MGC Library."

BASE COUNT 238 a 236 c 233 g 246 t

#### ORIGIN

Alignment Scores:  
 Pred. No.: 1..1e-130 Length: 953  
 Score: 1086.00 Matches: 198  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 100.00% Indels: 0  
 DB: 14 Gaps: 0

US-09-966-880A-8 (1-198) x B0065440 (1-953)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
 |||||||  
 Db 64 ATGGACAGCCTCTTGATGACCGGAGGAAGTTCTTACCAATTCACAAATGTCCGCTGG 123  
 |||||||  
 Qy 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40  
 |||||||  
 Db 124 GCTAAGGGTGGCGGTGAGACCTACTGTGTACGTAGTGAGAGCGGTGACAGTGCTACA 183  
 |||||||  
 Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
 |||||||  
 Db 184 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGATTCGTC 243  
 |||||||  
 Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
 |||||||  
 Db 244 TTCTCTCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGCTCACCTGG 303  
 |||||||  
 Qy 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
 |||||||  
 Db 304 TTCACCTCTGGAGCCCTCTGTACACTGTGCCCGACATGTGCCGACTTCTCGGAGGG 363  
 |||||||  
 Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
 |||||||  
 Db 364 AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 423  
 |||||||  
 Qy 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
 |||||||  
 Db 424 GCTAGAGCCGAGGGGCTGCGCGGCTGCACCGCGCGGTGCAATAGCCATCATGACC 483  
 |||||||  
 Qy 141 PheLysAspTyrPheTyrCysTyrAsnThrPheValGluAsnHisGluArgThrPheLys 160  
 |||||||  
 Db 484 TTCAAAGATTATTTTTACTGCTGGAATACTTTTGTAGAAACCAACCAAGAACTTTCAAA 543  
 |||||||  
 Qy 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
 |||||||  
 Db 544 GCCTGGGAAGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGCGCATCCTT 603  
 |||||||  
 Qy 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198  
 |||||||

Db 604 TTGCCCCGTATGAGGTTGATGACTTACGAGACGCAATTCGTACTTTGGGACTT 657

#### RESULT 5

B0055935

LOCUS

DEFINITION B0055935 1052 bp mRNA linear EST 29-MAR-2002  
 5', mRNA sequence.

ACCESSION B0055935

VERSION B0055935.1 GI:19815262

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 1052)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)

Tissue Procurement: Lou Staudt

cDNA Library Preparation: Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Agencourt Bioscience Corporation

Clone Distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM2051 row: m column: 14

High quality sequence stop: 665.

Location/Qualifiers

1..1052

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:5808181"

/clone\_lib="NIH\_MGC\_99"

/tissue\_type="lymphoma, cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: lymph; Vector: pOTB7; Site\_1: XhoI; Site\_2:

EcoRI; cDNA made by oligo-dt priming. Directionally cloned

into EcoRI/XhoI sites using the following 5' adaptor:

GGCAGGAG(G). Size-selected >500bp for average insert size

1.8kb. Library constructed by Ling Hong in the laboratory

of Gerald M. Rubin (University of California, Berkeley)

using ZAP-cDNA synthesis kit (Stratagene) and Superscript

II RT (Life Technologies). Note: this is a NIH\_MGC

Library."

BASE COUNT 276 a 250 c 258 g 266 t 2 others

ORIGIN

Alignment Scores:

Pred. No.: 1..31e-130 Length: 1052

Score: 1086.00 Matches: 198

Percent Similarity: 100.00% Conservative: 0

Best Local Similarity: 100.00% Mismatches: 0

Query Match: 100.00% Indels: 0

DB: 14 Gaps: 0

US-09-966-880A-8 (1-198) x B0055935 (1-1052)

Qy 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20

|||||

Db 64 ATGGACAGCCTCTTGATGACCGGAGGAAGTTCTTACCAATTCACAAATGTCCGCTGG 123

|||||

Qy 21 AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40

|||||

Db 124 GCTAAGGGTGGCGGTGAGACCTTACCTGTGTACGTAGTGAGAGCGGTGACAGTGCTACA 183

|||||

Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60

|||||

Db 184 TCCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCAGTGGATTCGTC 243

|||||

Qy 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80

|||||

```

Db      244  TTCCTCCGTACATCTCGGACTGGGACCTAGACCCCTGGCGCTGCTACCGCGTCACCTGG 303
QY      81  PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAlaAspPheLeuArgGly 100
Db      304  TTCACCTCTCGAGCCCTCTGCTACGACTGTGCCGACATGTGCCGACACTTTCGCCAGGG 363
QY      101  AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db      364  AACCCCAACCTCAGTCTGAGGATCTTCACGCGCGCTCTACTTCTGTGAGGACCGCAAG 423
QY      121  AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db      424  GCTGAGCCCGAGGGGCTGGCGGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 483
QY      141  PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db      484  TTCAAAGATATATTTTACTGCTGTAATCTTTGTAGAAACCATTGAAGAACTTTCANA 543
QY      161  AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db      544  GCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 603
QY      181  LeuProLeuTyrGluValAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db      604  TTGCCCTGTATGAGGTTGATGACTTACGAGACGCATTTCTGTACTTTGGGACTT 657

RESULT 6
BG757089
LOCUS   6027115124F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4855517 5',
DEFINITION
mRNA sequence.
ACCESSION
BG757089
VERSION
BG757089.1 GI:14067742
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 820)
AUTHORS
NIH-MGC http://mgi.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Prepared: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1704 row: o column: 06
High quality sequence stop: 675.
FEATURES
Location/Qualifiers
1..820
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4855517"
/clone_lib="NIH_MGC_48"
/tissue_type="primary B-cells from tonsils (cell line)"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
Site:2: EcoRI; cDNA made by oligo-dT priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by Ling
Hong in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
210 a 230 c 200 g 180 t
BASE COUNT
ORIGIN

```

## Alignment Scores:

Pred. No.: 1.17e-127 Length: 820  
 Score: 1062.00 Matches: 193  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 97.79% Indels: 0  
 DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x BG757089 (1-820)

```

QY      1  MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db      85  ATGCACAGCCTCTTGATGAACCGAGGAAGTTCTTTTACCATTTCAAAAATGTCGGCTGG 144
QY      21  AlalysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db      145  GCTAAGGGTCGGCTGAGACCTACCTGTCTAGTAGTAGAAGAGCGCTGACAGTGCTACA 204
QY      41  SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db      205  TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGAACGGCTGCCACGTGGAATTGCTC 264
QY      61  PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db      265  TTCTTCGGCTACATCTCGGACTCGGACCTAGACCTGGCGCTGCTACCGCGTCACTCTGG 324
QY      81  PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db      325  TTCACCTCTCGGAGCCCTCTGCTAGGACTGTGCCGACATGTGGCGGACTTTTCTCGGAGGG 384
QY      101  AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120
Db      385  AACCCCAACCTCAGTCTGAGGATCTTCACCGCGCGCTCTACTTCTGTGAGGACCGCAAG 444
QY      121  AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db      445  GCTGAGCCCGAGGGCTGCGCGGCTGCACCGCGCGGGTGCAATAGCCATCATGACC 504
QY      141  PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db      505  TTCAAAGATATTTTACTGCTGGAATCTTTGTAGAAACCATTGAAGAACTTTCANA 564
QY      161  AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db      565  GCCTGGGAAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTCGGCGCATCCTT 624
QY      181  LeuProLeuTyrGluValAspLeuArgAspAlaPhe 193
Db      625  TTGCCCTGTGAGGTTGATGACTTACGAGACGCATTTC 663

```

## RESULT 7

BG757392 602711022F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:4851580 5',  
 LOCUS DEFINITION  
 mRNA sequence.

ACCESSION BG757392  
 VERSION BG757392.1 GI:14068045  
 KEYWORDS EST.  
 SOURCE human.

## ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

## 1 (bases 1 to 693)

## AUTHORS

## TITLE

## JOURNAL

## COMMENT

NIH-MGC http://mgi.nci.nih.gov/.  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Prepared: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLCM1694 row: k column: 05  
 High quality sequence stop: 693.  
 Location/Qualifiers

# FEATURES

source  
 1..693  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4851580"  
 /clone\_lib="NIH\_MGC\_48"  
 /tissue\_type="primary B-cells from tonsils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: B-cells; Vector: pORF7; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."  
 BASE COUNT 157 a 188 c 178 g 170 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 1.78e-126 Length: 693  
 Score: 1052.00 Matches: 194  
 Percent Similarity: 98.48% Conservative: 0  
 Best Local Similarity: 98.48% Mismatches: 2  
 Query Match: 96.87% Indels: 1  
 DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x BG757392 (1-693)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
 DB 81 ATGGACAGCCTCTGTATGATACCGGAGGAGTCTTTACCAATTCAAATATGTCGGCTGG 140  
 QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLys-ArgArgAspSerAlaTh 40  
 DB 141 GCTAAGGGTGGGGTGAGAGCTACCTGTGTAGTAGTAACGAGGCGTGACAGTGCTAC 200  
 QY 40 rSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLe 60  
 DB 201 ATCTTTTCACTGGACTTTGGTATCTTCGCAATAAGACGGCTGCCACATGGGAATTGCT 260  
 QY 60 uPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTr 80  
 DB 261 CTTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGCTGCTACCGGCTCAGCTG 320  
 QY 80 pPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgL 100  
 DB 321 GTTCACTCTCTGGAGCCCTCTGCTACGACTGTGCCGACATGTGCCGACATTTCTGCGAG 380  
 QY 100 YAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLy 120  
 DB 381 GAACCCCAACCTCAGTCTGAGGATCTCACCGCGGCTCTACTTCTGTGAGGACGCCAA 440  
 QY 120 sAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140  
 DB 441 GGCTGAGCCGAGGGCTGGCGGCTGTGCACCGCGGGGTGCAATACCATCATGAC 500  
 QY 140 rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160  
 DB 501 CTTCAAGATATTCCTTACTGCTGGGAATATCTTTGTAGAAACCATGAAGAATCTTCAA 560  
 QY 160 sAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLe 180  
 DB 561 AGCCTGGGAGGGCTCATGAAATTCAGTTCGCTCTCCAGACACCTTCGGCGCATCCT 620  
 QY 180 uLeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrTr 196  
 DB 621 TTGCGCCCTGTATGAGTTGATGACTTAGGAGCGCATTTGCTACTTTG 669

## RESULT 8

BF975166  
 LOCUS 602244657F1 NIH\_MGC\_48 Homo sapiens cDNA clone IMAGE:433539 5',  
 DEFINITION mRNA sequence.  
 ACCESSION BF975166  
 VERSION BF975166  
 KEYWORDS BF975166.1 GI:12342381  
 SOURCE EST.  
 ORGANISM human.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 942)  
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D.  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLCM1207 row: a column: 16  
 High quality sequence stop: 707.  
 Location/Qualifiers  
 1..942  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:433539"  
 /clone\_lib="NIH\_MGC\_48"  
 /tissue\_type="primary B-cells from tonsils (cell line)"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: B-cells; Vector: pORF7; Site\_1: XhoI;  
 Site\_2: EcoRI; cDNA made by oligo-dT priming.  
 Directionally cloned into EcoRI/XhoI sites using the  
 following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
 for average insert size 1.8kb. Library constructed by Ling  
 Hong in the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies).  
 Note: this is a NIH\_MGC Library."  
 BASE COUNT 249 a 209 c 251 g 233 t  
 ORIGIN

Alignment Scores:  
 Pred. No.: 9.04e-124 Length: 942  
 Score: 1033.00 Matches: 194  
 Percent Similarity: 97.00% Conservative: 0  
 Best Local Similarity: 97.00% Mismatches: 4  
 Query Match: 95.12% Indels: 2  
 DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x BF975166 (1-942)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGln-PheLysAsnValArgTr 20  
 DB 78 ATGGACAGCCTCTTTGATGAACCGGAGGAGTCTTTTACCAGTTCAAAATATGTCGGCTG 137  
 QY 20 pAlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaTh 40  
 DB 138 GGCTAAGGGTGGGGTGAGAGCTACCTGTGTAGTAGTAAGAGGCGTGACAGTGCTAC 197  
 QY 40 rSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLe 60  
 DB 198 ATCTTTTCACTGGACTTTGGTATCTTCGCAATAAGACGGCTGCCACATGGGAATTGCT 257  
 QY 60 uPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTr 80  
 DB 258 CTTCTCCGCTACATCTCGGACTGGGACCTAGACCTGGCGGCTGCTACCGGCTCAGCTG 317



QY 80 pPhethrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgL 100  
|||||  
Db 318 GTTCACTCTCTGGAGCCCTGCTACGACTGTGCCGACATGTGGCCGACTTTCTGCGAGG 377  
QY 100 yAsnProAsnLeuSerLeuArgIlePhePheThrAlaArgLeuTyrPheCysGluAspArgLy 120  
|||||  
Db 378 GAACCCCAACCTCAGTCTGAGATCTTACCCGCGCCCTCTACTTCTGTGAGGACGCCAA 437  
QY 120 salaGluProGluGlyLeuArgLeuHisArgAlaGlyValGlnIleAlaIleMetTh 140  
|||||  
Db 438 GCTGAGCCCGAGGGCTGCGCGCTGCACGAGCCGGGTGCAATAGCCATCATGAC 497  
QY 140 rPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLy 160  
|||||  
Db 498 CTTCAAAGATTATTTTACTGCTGGAATACTTTGTAGAAACCACATGAAAGAACTTTCAA 557  
QY 160 salaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgIleLe 180  
|||||  
Db 558 AGCCTGGGAGGGCTGCATGAAATTCAGTTCTCTCCAGACAGCTTTCGGCGCATCCT 617  
QY 180 uLeuProLeuTyrGlu-ValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198  
|||||  
Db 618 TTGGGCCCTGTATGAGGGTTGATGACTTACGAGACCAATTCGTACTTGGGACTTG 673

RESULT 9  
AJ446140  
LOCUS 757 bp mRNA linear EST 19-APR-2002  
DEFINITION AJ446140 riken1 Gallus gallus CDNA clone 1219rl, mRNA sequence.  
ACCESSION AJ446140  
VERSION AJ446140.1 GI:20213361

KEYWORDS EST.  
SOURCE chicken.  
ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

REFERENCE 1 (bases 1 to 757)  
AUTHORS Buerstedde,J.M.  
TITLE Gallus gallus bursal lymphocyte EST  
JOURNAL Unpublished (2002)  
COMMENT Contact: Buerstedde JM  
Cellular Immunology  
Heinrich-Pette-Institute  
Martinistr. 52, 20251 Hamburg, Germany  
Email: URL: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.

FEATURES  
Location/Qualifiers  
1..757

source  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/clone="1219rl"  
/clone\_lib="riken1"  
/cell\_type="bursal lymphocyte"  
/dev\_stage="2-3 weeks old"  
/note="CB inbred strain"

BASE COUNT 180 a 205 c 184 g 188 t  
ORIGIN

Alignment Scores:  
Pred. No.: 8.58e-118 Length: 757  
Score: 986.00 Matches: 178  
Percent Similarity: 94.44% Conservative: 9  
Best Local Similarity: 89.90% Mismatches: 11  
Query Match: 90.73% Indels: 0  
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AJ446140 (1-757)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20  
|||||  
Db 97 ATGGACAGCCTCTTGATGAGAGGAGGCTCTTCTCAATTCAGAACCTTGGCTGG 156  
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40

Db 157 GCCAAAGGCGTGTGTAACCTACCTCTGTATGTTGTAAGCCCGTGACAGTGCTACA 216  
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60  
|||||  
Db 217 TCATGCTCCCTGACCTTTGGATACCTGCGTAAACAAGATGGTGGCCATGTGGAGGTCTC 276  
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80  
|||||  
Db 277 TTCCTACGCTACATCTCAGCTTGGACCTGGACCCAGCCGCTGCTACCCGCATCACATGG 336  
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100  
|||||  
Db 337 TTCACCTCTCTGGAGCCCTGTTATGACTGTGCCGACATGTGGCTGACTTCTCTCGTGCC 396  
QY 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120  
|||||  
Db 397 TACCAAACTTGACCTCCGCATTTTCTACCTGCCCGCTCTACTTCTGTGAAGATCGCAAG 456  
QY 121 AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140  
|||||  
Db 457 GCTGAGCCTCAGGGGCTGAGACGCTGCACCGGCTGGGGCCCAAAATCGCATCATGACT 516  
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160  
|||||  
Db 517 TTCAAAGATTCTTCTACTGCTGGAACACACTTGTGGAGAACAAGGAAAAAGACATTCAAA 576  
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180  
|||||  
Db 577 GCTGGGAAGGGCTGCATGAAACTCTGTCCATCTGTCCAGGAAACTCCGACGGATCCTT 636  
QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198  
|||||  
Db 637 CTGCCACTGTATGAAGTAGATGATTACGAGATGCCTTTAAAAACTCTGGGACTT 690

RESULT 10  
AJ453647  
LOCUS 696 bp mRNA linear EST 22-APR-2002  
DEFINITION AJ453647 riken1 Gallus gallus CDNA clone 35b23rl, mRNA sequence.  
ACCESSION AJ453647  
VERSION AJ453647.1 GI:20263743

KEYWORDS EST.  
SOURCE chicken.  
ORGANISM Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

REFERENCE 1 (bases 1 to 696)  
AUTHORS Buerstedde,J.M.  
TITLE Gallus gallus bursal lymphocyte EST  
JOURNAL Unpublished (2002)  
COMMENT Contact: Buerstedde JM  
Cellular Immunology  
Heinrich-Pette-Institute  
Martinistr. 52, 20251 Hamburg, Germany  
Email: URL: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.

FEATURES  
Location/Qualifiers  
1..696

source  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/clone="35b23rl"  
/clone\_lib="riken1"  
/cell\_type="bursal lymphocyte"  
/dev\_stage="2-3 weeks old"  
/note="CB inbred strain"

BASE COUNT 160 a 193 c 173 g 170 t  
ORIGIN

Alignment Scores:  
Pred. No.: 1.11e-116 Length: 696  
Score: 977.00 Matches: 177  
Percent Similarity: 93.94% Conservative: 9  
Best Local Similarity: 89.39% Mismatches: 12

```

Query Match: 89.96% Indels: 0
DB: 9 Gaps: 0
US-09-966-880A-8 (1-198) x AJ453647 (1-696)

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 97 ATGGACAGCCTCTTGATGAGAGGAAGCTCTTCTCTACAATTTCAAGAACCCTGCGCTGG 156
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 157 GCCAAAGGCGCTGCTGAACCTACCTCTGTTATGTTGTAAGCGCGTGACAGTGCTACA 216
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 217 TCATGCTCCCTGGACTTTGGATACCTGCGTAACAAGATGGGTGGCCATGTGGAGTTCTC 276
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 277 TTCCTACGCTACATCTTGTATGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 336
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 337 TTCACCTCTCGAGCGCTGATGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 396
QY 101 AsnProAsnLeuSerLeuArgGlyLeuArgHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 397 TACCAAACTTGACCTCGCATTTTCTACTGCGCCGCTTACTTCTGTGAAGATCGCAAG 456
QY 121 AlaGluProGluGlyLeuArgHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 457 GCTGAGCCTGAGGCGCTGAGACGCTGACCGCGCTGAGCGCGCTGAGCGCGCTGAGCG 516
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 517 TTCAAGATTTCTTCTACTGCTGAACACGTTTGTGGAGAACAGGAAAGACATTCAAA 576
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 577 GCCTGGGAAGGCGTGCATGAAACTCTGTCATCTGTCACAGAACTCTGAGGACCTT 636
QY 181 LeuProLeuTyrGluValAspAspLeuArgAspAlaPheArgThrLeuGlyLeu 198
Db 637 CTGCACCTGTATGAAGTAGATGATTTACGAGATGCCCTTTTAAACTCTGGGACTT 690

RESULT 11
AJ450296 729 bp mRNA linear EST 19-APR-2002
LOCUS riken1 Gallus gallus cdna clone 24m22r1, mRNA sequence.
DEFINITION
ACCESSION AJ450296
VERSION AJ450296.1 GI:20217517
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.
Buerstedde,J.M.
1 (bases 1 to 729)
Gallus gallus bursal lymphocyte EST
Unpublished (2002)
Contact: Buerstedde JM
Cellular Immunology
Heinrich-Pette-Institute
Martinistr. 52, 20251 Hamburg, Germany
Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
Location/Qualifiers
1. .729
/organism="Gallus gallus"
/db_xref="taxon:9031"
/clone="24m22r1"
/clone_lib="riken1"
/cell_type="bursal lymphocyte"

FEATURES
source

```

```

/dev_stage="2-3 weeks old"
/note="CB inbred strain"
BASE COUNT 173 a 199 c 187 g 170 t
ORIGIN

```

## Alignment Scores:

```

Pred. No.: 6,77e-108 Length: 729
Score: 910.00 Matches: 171
Percent Similarity: 92.75% Conservative: 8
Best Local Similarity: 88.60% Mismatches: 11
Query Match: 83.75% Indels: 3
DB: 9 Gaps: 0

```

US-09-966-880A-8 (1-198) x AJ450296 (1-729)

```

QY 1 MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp 20
Db 157 ATGGACAGCCTCTTGATGAGAGGAAGCTCTTCTCTACAATTTCAAGAACCCTGCGCTGG 216
QY 21 AlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThr 40
Db 217 GCCAAAGGCGCTGCTGAACCTACCTCTGTTATGTTGTAAGCGCGTGACAGTGCTACA 276
QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
Db 277 TCATGCTCCCTGGACTTTGGATACCTGCGTAACAAGATGGGTGGCCATGTGGAGTTCTC 336
QY 61 PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
Db 337 TTCCTACGCTACATCTTGTATGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 396
QY 81 PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
Db 397 TTCACCTCTCGAGCGCTGATGACCTGACCTGACCTGACCTGACCTGACCTGACCTGAC 456
QY 101 AsnProAsnLeuSerLeuArgGlyLeuArgHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 457 TACCAAACTTGACCTCGCATTTTCTACTGCGCCGCTTACTTCTGTGAAGATCGCAAG 516
QY 121 AlaGluProGluGlyLeuArgHisArgAlaGlyValGlnIleAlaIleMetThr 140
Db 517 GCTGAGCCTGAGGCGCTGAGACGCTGACCGCGCTGAGCGCGCTGAGCGCGCTGAGCG 576
QY 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160
Db 577 TTCAAGATTTCTTCTACTGCTGAACACGTTTGTGGAGAACAGGAAAGACATTCAAA 636
QY 161 AlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 637 GCCTGGGAAGGCGTGCATGAAACTCTGTCATCTGTCACAGAACTCTGAGGACCTT 696
QY 180 eu-LeuProLeuTyrGluValAspAspLeuArg 190
Db 697 TTCCTGCCACTGTATGAAGTAGATGATTACGA 729

```

## RESULT 12

AJ450317

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Chickens.

Gallus gallus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Archosauria; Aves; Neognathae; Galliformes; Phasianidae;

Phasianinae; Gallus.

1 (bases 1 to 688)

Buerstedde,J.M.

Gallus gallus bursal lymphocyte EST

Unpublished (2002)

Contact: Buerstedde JM

Cellular Immunology

Heinrich-Pette-Institute  
Martinistr. 52, 20251 Hamburg, Germany  
Email: URL: <http://genetics.hpi.uni-hamburg.de/dt40est.html>.

FEATURES	
source	Location/Qualifiers
1.	.688

```

source
1 1688
/organism="Gallus gallus"
/db_xrefs=taxon:9031
/clone="24n2rl"
/clone_lib="rikenl"
/cell_type="bursal lymphocyte"
/dev_stage="2-3 weeks old"
/notes="CB inbred strain"
163 a 191 c 178 q 156 t
BASE COUNT

```

Alignment Scores:		
Pred. No.:	1,528-104	Length:
Score:	884.00	Matches:
Percent Similarity:	93.7%	Conservative:
Best Local Similarity:	89.27%	Mismatches:
Query Match:	81.40%	Indels:
DB:	9	Gaps:
		0
		0
		688

US-09-966-880A-8 (1-198) x AJ450317 (1-688)

QY	1	MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrp	20
DB	157	ATGGACACCTCTTGATGAAGGAAGCTCTTCCTCTACAATTTCAAGAACCTCGCGTGG	216
QY	21	AlaLysGlyArgGluThrTyrLeuCysTyrValValLysArgAspSerAlaThr	40
DB	217	GCCAAAGGCGTCGTGAACCTACTCTGTGTATGTTGAAGCGCGTGACAGTGTCTACA	276

Qy 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60

[illegible]

Db 337 TTCTTACGCTACATCTCAGCTTGGGACCTGGACCCAGGCGGTGCTACGGCATCACATGG 396

81 PheHrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100

Qy 101 AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLys 120

Db 457 TACCCAAACTTGACCCCTCCGCATTTTCACTGCCCGGCTCTACTTCTGTGAAGATCGCAAG 516

[illegible]

Qy 141 PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLys 160

Db 577 TTCAAAGATTCTTCTACTGCTGGAACACGTTTGTGGAGACAGGGAAAGACATTCAA 636

Db 637 GCCTGGGAAGGCTGCATGAAACTCTGTCCATCTCTCAGGAAACTCCGA 687

RESULT 13  
BG144705

LOCUS	BGI14703	522 bp	miRNA	linear	EST V1-FEB-200
DEFINITION	U73f07.y1 Soares_mouse_NWBG_bcell Mus musculus cDNA clone				
IMAGE	322627	5'	similar to	NC_009580	COX2A2 ACNTMAM.ONI INDICED

ACCESSION  
BG144705  
CYTIDINE DEAMINASE. ; mRNA sequence.

**KEYWORDS** EST.  
**SOURCE** house mouse.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 522)
TITLE	NCI-CGAP <a href="http://www.ncbi.nlm.nih.gov/ncicgap">http://www.ncbi.nlm.nih.gov/ncicgap</a> . National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL	Unpublished (1997)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgaps-r@mail.nih.gov">cgaps-r@mail.nih.gov</a> This clone is available royalty-free through LLNL ; contact the IMAGE Consortium ( <a href="mailto:info@image.llnl.gov">info@image.llnl.gov</a> ) for further information. MGI:1077801

Seq primer: -40RP from Gibco  
High quality sequence stop: 487.

FEATURES  
SOURCE

1. .522 LOCATION/Qualifiers

```
/organism="Mus musculus"
/db_xref="taxon:10090"
```

/organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /clone\_lib="IMAGE:333637"  
 /clone\_lib="Soares\_mouse\_NMGb\_bcell"  
 /lab\_host="DH10B (phage-resistant)"  
 /notes="Organ: germinal B-cell; Vector: pT73D-Pac  
 (Pharmacia) with a modified polylinker; Site.1: Not I;  
 Site.2: Eco RI; 1st strand cDNA was primed with a Not I -  
 oligo(dT) primer [5'  
 TGTTACCAATCTGAAGTGGAGGCGCGCTGTTTTTTTTTTTTTTTTTTT  
 T 3']; double-stranded cDNA was ligated to Eco RI  
 adaptors (Pharmacia), digested with Not I and cloned into  
 the Not I and Eco RI sites of the modified pT73 vector.  
 Library is normalized; constructed by Bento Soares and  
 M. Fatima Ronaldo."

BASE COUNT	113 a	140 c	140 q	129 t
------------	-------	-------	-------	-------

Alignment Scores:

Argument Scores:					
Pred. No.:	1.18e-102	Length:	522		
Score:	868.00	Matches:	160		
Percent Similarity:	96.47%	Conservative:	47		
Best Local Similarity:	94.12%	Mismatches:	5		
Query Match:	79.93%	Indels:	1		
DB:	12	Gaps:	0		

US-09-966-880A-8 (1-198) x BGL44705 (1-522)

**Qy**      LeuCysTyr - ValValLysArgAspSerAlaThrSerPheSerLeuAspPheGlyTy 48  
          |||||  
          |||:

[illegible]

Db 61 CCTTCGCAACAAGTCTGGCTGCCACCGTGGAAATTGTTTCTACGCTACATCTCAGACTG 120

QY 68 pAspLeuAspProGlyArgCysTyrArgValThrTrpPheHrSerTrpSerProCysTyr 88

Qy 88 rAspCysAlaArqHisValAlaAlaSpPheLeuArgGlyAsnProAsnLeuSerLeuArgII 108

D6 181 TGACTGTGCCGGCACGTGGCTGAGTTTCTGAGATGGAACCCCTAACCTCAGCCTGAGGAT 240

QY	100	erfneimnrmnlgvbeuifrfnefvsolunspnlgufyfnsladgufurloagwufyfeunfynl	142
Db	241	TTTTACCGCGGGCCCTACTCTCTGTGAAGACCGCAAGSCTGASGCTGASGGCTGCGGAG	300

Qy 128 gLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyrCysTr 148

301 ACTGCACCGCGCTGGGGTCCAGATCGGGGATCATGACCTTCAAGAGACTATTTTACTCTG 360

Db 361 GAATACATTTCTAGAAAATCGTGAAGAAGACTTTCAAAGCCTGGGAAGGCTACATGAAA 420

Qy 168 nSerValArgLeuSerArgGlnLeuArgArgIleLeuLeuProLeuTyrGluValAspAs 188

```

Db      421  TTCTGTCGGGTAAACAGACAACTTCGGCGCATCTTTTGCCTTGTTACGAAGTCGATGA 480
Qy      188  pLeuArgAspAlaPheArgThrLeuGly 197
          |||||
Db      481  CTTGCGAGATGTCATTTGCTATGTGGGA 508

RESULT 14
LOCUS   BF238155
DEFINITION 601811880F1 NIH_MGC_48 Homo sapiens cDNA clone IMAGE:4054915 5',
          mRNA sequence.
ACCESSION BF238155
VERSION   BF238155.1 GI:11152074
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 541)
           NIH-MGC http://mgi.nci.nih.gov/.
           National Institutes of Health, Mammalian Gene Collection (MGC)
           Unpublished (1999)
           Contact: Robert Strausberg, Ph.D.
           Email: cgabbs@mail.nih.gov
           Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
           cDNA Library Preparation: Ling Hong/Rubin Laboratory
           DNA Library Arrayed by: The I.M.A.G.E. Consortium (ILLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
           found through the I.M.A.G.E. Consortium/ILLNL at:
           http://image.llnl.gov
           Plate: LLCW895 row: p column: 20
           High quality sequence stop: 541.
           Location/Qualifiers
             1..541
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:4054915"
               /clone_lib="NIH_MGC_48"
               /tissue_type="primary B-cells from tonsils (cell line)"
               /lab_host="DH10B (phage-resistant)"
               /note="Organ: B-cells; Vector: pOTB7; Site:1: XhoI;
               Site:2: EcoRI; cDNA made by oligo-dT priming.
               Directionally cloned into EcoRI/XhoI sites using the
               following 5' adaptor: GGCAGAG(G). Size-selected
               for average insert size 1.8kb. Library constructed by Ling
               Hong in the laboratory of Gerald M. Rubin (University of
               California, Berkeley) using ZAP-cDNA synthesis kit
               (Stratagene) and Superscript II RT (Life Technologies).
               Note: this is a NIH_MGC Library."
           BASE COUNT 120 a 150 c 142 g 129 t
           ORIGIN

Alignment Scores:
Pred. No.: 2,53e-101 Length: 541
Score: 858.00 Matches: 154
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 79.01% Indels: 0
DB: 12 Gaps: 0

US-09-966-880A-8 (1-198) x BF238155 (1-541)

Qy      1  MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrglnPheLysAsnValArgTrp 20
          |||||
Db      78  ATGGACAGCCTCTTGATGAACCGGAGGAGTCTTCTTACCAATTCAAAATGTCGCTGG 137

Qy      21  AlaLysGlyArgArgGluThrTyrlLeuCysTyrlValValLysArgArgAspSerAlaThr 40
          |||||
Db      138  GCTAAGGGTCGGGTGAGACCTACCTTGCTACGTAGTAGAAGAGCGGTGACAGTCTCTACA 197

Qy      41  SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60
          |||||

```

```

Db      198  TCCTTTTCACTGGACTTTGGTTATCTTCGCAATAAGACGGCTGCCACGTGAATTTGCTC 257
Qy      61  PheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrp 80
          |||||
Db      258  TTCTCTCCGCTACATCTCGGACTGGGACCTAGACCCCTGGCCGCTGTACCGCTGACCTGG 317
Qy      81  PheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGly 100
          |||||
Db      318  TTACACCTCTCGAGGCCCTGCTACGACTGTGCCCGACATGTGCCGCACTTCTTGGAGGG 377
Qy      101  AsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrlPheCysGluAspArgLys 120
          |||||
Db      378  AACCCCAACCTCAGTCTGAGGATCTCACCGCGCCCTCTACTTCTGTGAGACGCGCAAG 437
Qy      121  AlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThr 140
          |||||
Db      438  GCTGAGCCGAGGGGCTGCGGGCTGCACCGCGGGGTGCAATAGCCATCATGACC 497
Qy      141  PheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsn 154
          |||||
Db      498  TTCAAAGATTATTTTACTGCTGGAATACTTTTGTAGAAAAC 539

RESULT 15
LOCUS   AJ449745
DEFINITION 650 bp mRNA linear EST 19-APR-2002
ACCESSION AJ449745
VERSION   AJ449745.1 GI:20216966
KEYWORDS  EST.
SOURCE    chicken.
ORGANISM  Gallus gallus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
           Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 650)
           Buerstedde,J.M.
           Gallus gallus bursal lymphocyte EST
           Unpublished (2002)
           Contact: Buerstedde JM
           Cellular Immunology
           Heinrich-Pette-Institute
           Martinstr. 52, 20251 Hamburg, Germany
           Email: URL: http://genetics.hpi.uni-hamburg.de/dt40est.html.
           Location/Qualifiers
             1..650
               /organism="Gallus gallus"
               /db_xref="taxon:9031"
               /clone="23a2r1"
               /clone_lib="riken1"
               /cell_type="bursal lymphocyte"
               /dev_stage="2-3 weeks old"
               /note="CB inbred strain"
           BASE COUNT 150 a 182 c 167 g 151 t
           ORIGIN

Alignment Scores:
Pred. No.: 3.82e-100 Length: 650
Score: 850.00 Matches: 156
Percent Similarity: 93.22% Conservative: 9
Best Local Similarity: 88.14% Mismatches: 12
Query Match: 78.27% Indels: 1
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x AJ449745 (1-650)

Qy      1  MetAspSerLeuLeuMetAsnArgArgLysPheLeuTyrglnPheLysAsnValArgTrp 20
          |||||
Db      120  ATGGACAGCCTCTTGATGAACCGGAGGAGTCTTCTTCAAAATTCAAAATGTCGCTGG 179

Qy      21  AlaLysGlyArgArgGluThrTyrlLeuCysTyrlValValLysArgArgAspSerAlaThr 40
          |||||
Db      180  GCCAAGGCGCTCGTGAACCTACCTCTGTATCTTGTGTGAAGCCCGCTGACAGTCTACA 239

```





GenCore version 5.1.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model  
Run on: June 19, 2003, 17:42:06 ; Search time 51 Seconds  
(without alignments)  
1190.627 Million cell updates/sec

Title: US-09-966-880A-8  
Perfect score: 1086  
Sequence: 1 MDSLNNRRKFLYQFKVNRW.....ILLPLYEVDDLRFRTLGL 198

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0  
Searched: 441362 seqs, 153338381 residues  
Total number of hits satisfying chosen parameters: 882724  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+p2n.model -DEV=xlh  
-Q=/cgn2\_1/USFT0.spool/US09966880/runat\_14062003\_175646\_16331/app\_query.fasta\_1.391  
-DB=Issued\_Patents.NA -OFT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cd1  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -NAXLEN=2000000000  
-USER=US09966880 @CNU\_1.1.40 @runat\_14062003\_175646\_16331 -NCPU=6 -ICPU=3  
-NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents.NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5b\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	369.5	34.0	610	1	US-08-816-241-2
2	369.5	34.0	610	3	US-09-128-395-2
3	271	25.0	891	1	US-08-687-895-2
4	271	25.0	891	2	US-09-040-482-2
5	210	19.3	879	1	US-08-158-682A-3
6	201	18.5	879	1	US-08-158-682A-1
7	201	18.5	879	1	US-08-015-203-1
8	87.5	8.1	1806	3	US-09-068-655-10
9	87.5	8.1	4713	3	US-09-068-655-3
10	80	7.4	7077	4	US-09-221-017B-123
11	77.5	7.1	1209	6	5352575-4
12	75.5	7.0	3043	4	US-09-049-698-16

C	13	75.5	7.0	3181	4	US-09-049-698-18	Sequence 18, Appl
	14	73.5	6.8	2629	1	US-08-200-807-1	Sequence 1, Appli
	15	73.5	6.8	2629	1	US-08-488-305A-1	Sequence 1, Appli
C	16	72.5	6.7	9179	4	US-09-453-702B-162	Sequence 162, App
	17	72.5	6.7	13188	4	US-08-961-527-70	Sequence 70, Appl
C	18	72	6.6	1672	4	US-09-495-066-1	Sequence 1, Appli
	19	72	6.6	4201	1	US-08-080-255-4	Sequence 4, Appli
	20	72	6.6	4201	3	US-08-465-713-4	Sequence 4, Appli
	21	72	6.6	4201	5	PCT-US93-05857-4	Sequence 4, Appli
	22	72	6.6	11907	4	US-08-061-376-4	Sequence 4, Appli
	23	72	6.6	14255	1	US-08-320-559-1	Sequence 1, Appli
	24	72	6.6	14255	1	US-08-327-392-1	Sequence 1, Appli
	25	72	6.6	14255	1	US-08-306-691B-55	Sequence 55, Appl
	26	72	6.6	14255	3	US-08-545-860D-1	Sequence 1, Appli
	27	72	6.6	14255	5	PCT-US94-04496-1	Sequence 1, Appli
	28	71	6.5	2534	5	PCT-US93-06251-10	Sequence 10, Appl
C	29	71	6.5	8220	2	US-08-568-459A-11	Sequence 11, Appl
	30	71	6.5	8220	2	US-08-487-826B-11	Sequence 11, Appl
	31	71	6.5	8220	4	US-09-210-288-11	Sequence 11, Appl
C	32	71	6.5	19124	2	US-08-487-826B-13	Sequence 13, Appl
C	33	70.5	6.5	704	4	US-08-998-416-977	Sequence 977, App
C	34	70	6.4	2156	2	US-08-899-514-1	Sequence 1, Appli
	35	70	6.4	2254	4	US-08-552-369-1	Sequence 1, Appli
	36	70	6.4	7616	5	PCT-US94-01149-54	Sequence 54, Appl
	37	69.5	6.4	1724	4	US-09-385-259-1	Sequence 1, Appli
	38	69.5	6.4	1724	4	US-09-645-370-1	Sequence 1, Appli
	39	69.5	6.4	6496	4	US-09-221-017B-543	Sequence 543, App
C	40	69	6.4	1035	4	US-08-191-160-4	Sequence 4, Appli
C	41	69	6.4	1770	4	US-08-191-160-3	Sequence 3, Appli
	42	69	6.4	7003	2	US-08-378-548-24	Sequence 24, Appl
C	43	69	6.4	8224	6	5180808-1	Patent No. 5180808
C	44	69	6.4	4403765	4	US-09-103-840A-2	Sequence 2, Appli
C	45	69	6.4	4411529	4	US-09-103-840A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-816-241-2  
; Sequence 2, Application US/08816241  
; Patent No. 5804185  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL RNA EDITING ENZYME  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/816,241  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0239 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 2:





Db 520 TTAATAAACCAACTTTTCGACTTCTGAAAGAGAGCTACGGGAGAGTCTC 567

## RESULT 3

US-08-687-895-2  
; Sequence 2, Application US/08687895  
; Patent No. 5747319  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICANT: Au-Young, Janice  
APPLICANT: Hillman, Jennifer L.  
FILING DATE: Filed Herewith  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0109 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 891 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
LIBRARY: MUSCNOT1  
CLONE: 57953

US-08-687-895-2

Alignment Scores:  
Pred. No.: 3.73e-25 Length: 891  
Score: 271.00 Matches: 63  
Percent Similarity: 50.00% Conservative: 25  
Best Local Similarity: 35.80% Mismatches: 76  
Query Match: 24.95% Indels: 12  
DB: 1 Gaps: 5

US-09-966-880A-8 (1-198) x US-08-687-895-2 (1-891)

QY 11 PheLeuTyrGlnPheLysAsnValArgTrrAlaLysGlyArgArgGluThrTyrLeuCys 30  
Db 306 TTTAATTCAGTTCGGGAATGGAGTGGAGTTCGGGAGGAAACACCTTCCTCTGC 365  
QY 31 TyrValValLysArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyrLeuArg 50  
Db 366 TATGTGTTGAGCAGCAGGCGGAGGGGCGGAGTCCAGGCATCTCGGGGATACCTAGAG 425  
QY 51 AsnLysAsn--GlyCysHisValGluLeuLeuPheLeuArgTyrTrpSerAspTrrAsp 69  
Db 426 GATGAGCATCGGCTGCCCATGCAGAGGAAGCTTTCTTCAACACCATCTGCGCA---GCC 482  
QY 70 LeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAsp 89  
Db 483 TTCGACCCAGCCCTCGGGTACAAATGTCACCTGGTATGTCTCTCCAGCCCTGTGAGCG 542  
QY 90 CysAlaAlaHisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePhe 109

Db 543 TGTGCTGACCGCATTAACAAACCTTAGCAAGACCAAGACCTGCTGCTCATTCCTG 602  
QY 110 ThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu-----GlyLeu 126  
Db 603 GTGGGTGCGACTCTTCATGTGGGAG-----GAGCCGGAGATCCAGGCTGCTCTG 650  
QY 127 ArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyr 146  
Db 651 AAGAAGCTGAAGGAGCTGGCTCTAAACTCCGCATCATGAAGCCCGCAGGACTTCGAATAT 710  
QY 147 CysTrpAsnThrPheValGluAsnHisGlu-----ArgThrPheLysAlaTrpGlu 163  
Db 711 GTCTGGCAGAATTTCTGGAGCAAGAAGAGGGTGAATCCAAAGGCCTTTCARCCCTGGGAG 770  
QY 164 GlyLeuHisGluAsnSerValArgLeuSerArgGluLeuArgIle 179  
Db 771 GACATTGAGAGAACTTCCTATATACAGGAGGAAGTTGGCAGACATC 818

## RESULT 4

US-09-040-482-2  
; Sequence 2, Application US/09040482  
; Patent No. 5916556  
; GENERAL INFORMATION:  
; APPLICANT: Au-Young, Janice  
; APPLICANT: Hillman, Phillip R.  
; APPLICANT: Hillman, Jennifer L.  
; TITLE OF INVENTION: A NOVEL HUMAN MRNA EDITING ENZYME  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: U.S.  
; ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 1.5  
CURRENT APPLICATION DATA:  
APPLICANT: Au-Young, Janice  
APPLICANT: Hillman, Jennifer L.  
FILING DATE: US/09/040,482  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/687,895  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0109 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 891 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
LIBRARY: MUSCNOT1  
CLONE: 57953

US-09-040-482-2  
Alignment Scores:  
Pred. No.: 3.73e-25 Length: 891  
Score: 271.00 Matches: 63  
Percent Similarity: 50.00% Conservative: 25  
Best Local Similarity: 35.80% Mismatches: 76  
Query Match: 24.95% Indels: 12  
DB: 2 Gaps: 5

```
US-09-966-880A-8 (1-198) x US-09-040-482-2 (1-891)
Qy 11 PheLeuTyrGlnPheLysAsnValArgTirpAlaLysGlyArgGluThrTyrLeuCys 30
Db 306 TTTAAATCCAGTTCGGATGTGGATACAGTTCGGGAGAACACAGACCTTCTCTGC 365
Qy 31 TyrValValLysArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyrLeuArg 50
Db 366 TATGTGTTGAAGCACACAGGGCAAGGGGGCCAAAGTCAGGCATCTCGGGGATACCTAGAG 425
Qy 51 AsnLysAsn---GlyCysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAsp 69
Db 426 GATGAGCATGGCGTGGCCATGACAGAGAAAGCTTTCTCAACACCATCTCTGCCA---GCC 482
Qy 70 LeuAspProGlyArgCysTyrArgValThrTirpPheThrSerTrpSerProCysTyrAsp 89
Db 483 TTCGACCACGCCCTGCGGTACATGTCACTGTGTGTCTCCTCAGCCCTGTGCACGG 542
Qy 90 CysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePhe 109
Db 543 TGTGCTGACCGCATTTACAAACCCCTTAGCAAGACCAAGAACCTGCGTCTCTCATCTG 602
Qy 110 ThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu-----GlyLeu 126
Db 603 GTGGTGCATCTTCATGTGGGAG-----GAGCGGAGATCCAGGCTGCTCTG 650
Qy 127 ArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyr 146
Db 651 AAGAAGCTGAAGGAGGCTGCTGTAACTCGGCATCATGAAGCCCAAGGCTTCGATAT 710
Qy 147 CysTrpAsnThrPheValGluAsnHisGlu-----ArgThrPheLysAlaTrpGlu 163
Db 711 GTCTGGCAGAAATTTGTGGAGCAAGAGGGTGAATCCCAAGGCGTTTCARCCCTGGGAG 770
Qy 164 GlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIle 179
Db 771 GACATTCAGGAGAACTTCCCTATCTACTAGCAGGAGAAAGTTGGCAGACATC 818

RESULT 5
US-08-158-682A-3
; Sequence 3, Application US/08158682A
; Patent No. 5434058
; GENERAL INFORMATION:
; APPLICANT: Davidson, Nicholas O.
; TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: 321 No. 5434058th Clark Street, Suite 800
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158,682A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coolley, Ronald B.
; REGISTRATION NUMBER: 27,187
; REFERENCE/DOCKET NUMBER: ARCD:085
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 744-0090
; TELEFAX: (312) 245-4961
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 879 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 23..730
US-08-158-682A-3
Alignment Scores:
Pred. No.: 2,15e-17 Length: 879
Score: 210.00 Matches: 48
Percent Similarity: 56.30% Conservative: 28
Best Local Similarity: 35.56% Mismatches: 43
Query Match: 19.34% Indels: 16
DB: 1 Gaps: 5
US-09-966-880A-8 (1-198) x US-08-158-682A-3 (1-879)
Qy 24 ArgArgGluThrTyrLeuCysTyrValValLys-----ArgArgAspSerAla 39
Db 119 CGTAAAGAGGCGCTGCTGCTCTACGAAATCAAGTGGGCATGAGCCGGAAGATCTGGCGA 178
Qy 40 ThrSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCys---HisValGlu 58
Db 179 ACGTCA-----GGCAAAAACACCAACCAATCACGTGCA 211
Qy 59 LeuLeuPheLeu---ArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArg 77
Db 212 GTTAAATTTATAAAAAAATTTACGTCAGAAAGAGATTTTACCCATCATCAGCTGCACC 271
Qy 78 ValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPhe 97
Db 272 ATCAGCTGGTCTTCTGCTGGAGTCCCTGCTGGGAATGCTCCAGGCTATTAGAGAGTTT 331
Qy 98 LeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGlu 117
Db 332 CTGAGTCGGCACCCCTGGTGTGACTAGTATCTACGTAGCTCGGCTTTTGTGGCAGCATG 391
Qy 118 AspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAla 137
Db 392 GATCAACAA---AATCGCAAGGTCTCAGGACCTTGTTAACAGTGGAGTAATTTCAG 448
Qy 138 IleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheVal 152
Db 449 ATTATGAGAGCATCAGAGTATTATCACTGCTGGAGGAATTTTGTTC 493

RESULT 6
US-08-158-682A-1
; Sequence 1, Application US/08158682A
; Patent No. 5434058
; GENERAL INFORMATION:
; APPLICANT: Davidson, Nicholas O.
; TITLE OF INVENTION: Apolipoprotein B RNA Editing Protein:
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ARNOLD, WHITE & DURKEE
; STREET: 321 No. 5434058th Clark Street, Suite 800
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60610
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/158,682A
; FILING DATE:
; CLASSIFICATION: 435
```



QY 175 GlnLeuArgArgIleLeuLeuProLeuTyr 184  
 Db 572 -----TACGTACTGGAACCTCTAC 589

## RESULT 8

US-09-068-655-10/c  
 ; Sequence 10, Application US/09068655A  
 ; Patent No. 6136579  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACKSON, Ronald James  
 ; TITLE OF INVENTION: METHOD OF PRODUCING ALPHA 2,3-SIALYLTRANSFERASE  
 ; FILE REFERENCE: 026579-186  
 ; CURRENT APPLICATION NUMBER: US/09/068,655A  
 ; CURRENT FILING DATE: 1998-08-31  
 ; EARLIER APPLICATION NUMBER: AU PN 6587  
 ; EARLIER FILING DATE: 1995-11-15  
 ; EARLIER APPLICATION NUMBER: PCT/AU96/00725  
 ; EARLIER FILING DATE: 1996-11-15  
 ; NUMBER OF SEQ ID NOS: 13  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 10  
 ; LENGTH: 1806  
 ; TYPE: DNA  
 ; ORGANISM: Myxoma virus  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (2)..(439)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1405)..(1806)  
 ; US-09-068-655-10

Alignment Scores:  
 Pred. No.: 0.252 Length: 1806  
 Score: 87.50 Matches: 38  
 Percent Similarity: 41.07% Conservative: 8  
 Best Local Similarity: 33.93% Mismatches: 43  
 Query Match: 8.06% Indels: 23  
 DB: 3 Gaps: 7

US-09-966-880A-8 (1-198) x US-09-068-655-10 (1-1806)

QY 20 TrpAlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAla 39  
 Db 888 TGGTTCGGTTCAGTCCGCGACTTATCTGGCTATATACATGCTTACGAACAACGCAC 829  
 QY 40 ThrSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 59  
 Db 828 GCGTCGACCTGTCTAGATTGTGGAAGAACCGCTCCGACGTGGCGCGCGCTCCGCGGT 769  
 QY 60 LeuPheLeuArgTyrIleSerAsp-----TrpAspLeuAsp 71  
 Db 768 TCGTACGAATCCTACATCCGAAATATACGTACGATCGCGGTTCGCTATACAGACAACGT 709  
 QY 72 ProGlyArgCysTyrArgValThrTrpPheThrSerTrpSer-----ProCysTyr 88  
 Db 708 CCGGACGTTGACAGTTCCTCCATCGATGG-----GAATGGTTCGCGCTCGTACACCGCGTTAC 655  
 QY 89 AspCysAlaArgHisValAlaAspPheLeuArg-----GlyAsnPro----- 102  
 Db 654 ACGTGTGCCAGG-----GCGTAACGATTACGGGATTCCGGTACCCCAATGGGGGCCTA 601  
 QY 103 ----AsnLeuSerLeuArgIlePheThrAlaArgLeu 113  
 Db 600 ACCAATTATTCATT---ATTACACGGCTATACTA 568

## RESULT 9

US-09-068-655-3/c  
 ; Sequence 3, Application US/09068655A  
 ; Patent No. 6136579  
 ; GENERAL INFORMATION:  
 ; APPLICANT: JACKSON, Ronald James

; TITLE OF INVENTION: METHOD OF PRODUCING ALPHA 2,3-SIALYLTRANSFERASE  
 ; FILE REFERENCE: 026579-186  
 ; CURRENT APPLICATION NUMBER: US/09/068,655A  
 ; CURRENT FILING DATE: 1998-08-31  
 ; EARLIER APPLICATION NUMBER: AU PN 6587  
 ; EARLIER FILING DATE: 1995-11-15  
 ; EARLIER APPLICATION NUMBER: PCT/AU96/00725  
 ; EARLIER FILING DATE: 1996-11-15  
 ; NUMBER OF SEQ ID NOS: 13  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 3  
 ; LENGTH: 4713  
 ; TYPE: DNA  
 ; ORGANISM: Myxoma virus  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(549)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (1515)..(2081)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (2084)..(3745)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (3778)..(4434)  
 ; FEATURE:  
 ; NAME/KEY: CDS  
 ; LOCATION: (4440)..(4712)  
 ; US-09-068-655-3

Alignment Scores:  
 Pred. No.: 1.06 Length: 4713  
 Score: 87.50 Matches: 38  
 Percent Similarity: 41.07% Conservative: 8  
 Best Local Similarity: 33.93% Mismatches: 43  
 Query Match: 8.06% Indels: 23  
 DB: 3 Gaps: 7

US-09-966-880A-8 (1-198) x US-09-068-655-3 (1-4713)

QY 20 TrpAlaLysGlyArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAla 39  
 Db 998 TGGTTCGGTTCAGTCCGCGACTTATCTGGCTATATACATGCTTACGAACAACGCAC 939  
 QY 40 ThrSerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeu 59  
 Db 938 GCGTCGACCTGTCTAGATTGTGGAAGAACCGCTCCGACGTGGCGCGCGCTCCGCGGT 879  
 QY 60 LeuPheLeuArgTyrIleSerAsp-----TrpAspLeuAsp 71  
 Db 878 TCGTACGAATCCTACATCCGAAATATACGTACGATCGCGGTTCGCTATACAGACAACGT 819  
 QY 72 ProGlyArgCysTyrArgValThrTrpPheThrSerTrpSer-----ProCysTyr 88  
 Db 818 CCGGACGTTGACAGTTCCTCCATCGATGG-----GAATGGTTCGCGCTCGTACACCGCGTTAC 765  
 QY 89 AspCysAlaArgHisValAlaAspPheLeuArg-----GlyAsnPro----- 102  
 Db 764 ACGTGTGCCAGG-----GCGTAACGATTACGGGATTCCGGTACCCCAATGGGGGCCTA 711  
 QY 103 ----AsnLeuSerLeuArgIlePheThrAlaArgLeu 113  
 Db 710 ACCAATTATTCATT---ATTACACGGCTATACTA 678

## RESULT 10

US-09-221-017B-123  
 ; Sequence 123, Application US/09221017B  
 ; Patent No. 644799  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ROSS, Bruce C.  
 ; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF  
 ; NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:  
ADDRESSEE: MORRISON & FOERSTER  
STREET: 755 PAGE MILL ROAD  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304-1018

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows  
SOFTWARE: FastSeq for Windows Version 2.0b  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/221,017B  
FILING DATE: 23-DEC-1998

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1182

FILING DATE: 31-DEC-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP1546

FILING DATE: 30-JAN-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PP2911

FILING DATE: 09-APR-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/AU98/01023

FILING DATE: 10-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Monroy, Gladys H

REGISTRATION NUMBER: 32,430

REFERENCE/DOCKET NUMBER: 27340-20021.00

TELEPHONE: 650-813-5600

TELEFAX: 650-494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 123:

SEQUENCE CHARACTERISTICS:

LENGTH: 7077 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: UNKNOWN

ORGANISM: PORYPHYROMONAS GINGIVALIS

FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1...7077

US-09-221-017B-123

Alignment Scores:

Pred. No.: 17.5

Score: 80.00

Percent Similarity: 70.77

Best Local Similarity: 22

Query Match: 17

Indels: 16

Gaps: 3

US-09-966-880A-8 (1-198) x US-09-221-017B-123 (1-7077)

QY 4 LeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGly 23

Db 3900 TTGTATATAAATCGAAACGTTTATATATAGATCTCAGCCGTTTCGATGGGAAAGAC 3959

QY 24 ArgArgGluThrTyrLeuCysTyrValValLys-----ArgArgAspSerAlaThr 40

Db 3960 GCCGAGAAACATTTCTTGAACCATATAAAACCAAGGCACCCGCCGAGATTCCTGT--- 4016

QY 41 SerPheSerLeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeu 60

Db 4017 -----CGAACTCCGAGCGGGTGC-----CTTGCT 4040

QY 61 PheLeuArg 63

Db 4041 TTTTGTCTG 4049

RESULT 11

5352575-4

;Patent No. 5352575

; APPLICANT: PETROVSKIS, ERIK A.;POST, LEONARD E.;TIMMINS, JAMES G.

; TITLE OF INVENTION: PSEUDORABIES VIRUS PROTEIN

; NUMBER OF SEQUENCES: 12

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/513,282

; FILING DATE: 20-APR-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 100,817

; FILING DATE: 29-JUN-1987

; APPLICATION NUMBER: 886,260

; FILING DATE: 16-JUL-1986

; APPLICATION NUMBER: 784,787

; FILING DATE: 04-OCT-1985

; APPLICATION NUMBER: 801,799

; FILING DATE: 26-NOV-1985

; APPLICATION NUMBER: 844,113

; FILING DATE: 26-MAR-1986

;SEQ ID NO:4:

;LENGTH: 1209

5352575-4

Alignment Scores:

Pred. No.: 2.6

Score: 77.50

Percent Similarity: 34.91%

Best Local Similarity: 24.85%

Query Match: 7.14%

Indels: 6

Gaps: 6

US-09-966-880A-8 (1-198) x 5352575-4 (1-1209)

QY 24 ArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSer 43

Db 262 CGCGCGCCACGACGTACCGCGCCACGTG----- 288

QY 44 LeuAspPheGlyTyrLeuArgAsnLysAsnGlyCysHisValGluLeuLeuPheLeuArg 63

Db 289 -----GCCTGGTACCGCATCGCGGCGGGTGGCCACACCTGCTGTTATTCGAG 339

QY 64 TyrIleSerAspTrpAspLeuAspPro-----GlyArgCysTyrArgValThr 79

Db 340 TAGGCC-----GACTCGACCCCGACGAGGCTCTTGGGCGCTGCGCGCGCGCAC 390

QY 80 TrpPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArg 99

Db 391 ACGCCGATGTGGTGACCCCGCTCGCGGACTACATGTCCCGCGGAGGAGCGAGCTG--- 447

QY 100 GlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArg 119

Db 448 -----GGGCTGCTCATGTTGGCCCGCGGGGG 474

QY 120 LysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAla----- 137

Db 475 TTCAACGAGCGGCAGTACCGCGCTGTCGCTCGACGCGGTGAACATCTCACCAC 534

QY 138 ---IleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHisGlu 156

Db 535 TTGATGTCGTCGCTCCCGCGGGGCAAGAGTCCCGTTCGCCCGCTGGACGACAC--- 591

QY 157 ArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGlnLeu 176

Db 592 CGCACGTACAGTTCGCGCGCTGCTGGAGCGACGACAGCTCAAGCGGGCGGTGACGCTG 651

QY 177 ArgArgIleLeuLeuProLeuTyrGlu 185

```
Db      652 ATGCCATTCTGACGCCGCTTCTACCAG 678
RESULT 12
; Sequence 16, Application US/09049698
; Patent No. 6368792
; GENERAL INFORMATION:
; APPLICANT: BILLING-MEDEL, PATRICIA A.
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: HAYDEN, MARK
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: ROBERTS-RAPP, LISA
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STROUPE, STEPHEN D.
; TITLE OF INVENTION: REAGENTS AND METHODS FOR THE
; TITLE OF INVENTION: USEFUL FOR DETECTING DISEASES OF THE GASTROINTESTINAL
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/049,698
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,856
; FILING DATE: 31-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6068.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3043 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-049-698-16
Alignment Scores:
Pred. No.: 18 6 Length: 3043
Score: 75.50 Matches: 32
Percent Similarity: 43.09% Conservative: 21
Best Local Similarity: 26.02% Mismatches: 49
Query Match: 6.95% Indels: 21
DB: Gaps: 5
US-09-966-880A-8 (1-198) x US-09-049-698-16 (1-3043)
QY      58 GluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCys----- 75
Db      376 GAACCTGGTGGTGTATGGTTCATCTCTACCTGGAG-----TGTAGGTGG 332
QY      76 -----TyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHis 93
Db      331 TGCACACTATAACATCAGCATGTTATGGTTTTCANGCTTTTGGCCCTTTTGTACTGAGGATT 272
QY      94 ValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeu 113
Db      271 TTCCTTCCAAATCTCAGG-----AATTATATAGATACATTTTTCAAAAAAATCT 221
QY      114 TyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGly 133
Db      220 TTTTCTGT-----GGCTTCAACACAGGTAGCTAGAGCTCT 185
QY      134 ValGlnIleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheValGlu 153
Db      184 AGTCACCATATCCTCTATTTGTTCAATATTATTTTTCATCTTTCGCACACATAGGATCTAT 125
QY      154 AsnHisGluArgThrPheLysAla---TrpGluGlyLeuHisGluAsnSerValArgLeu 172
Db      124 AACATGACATATCTTCAAGCCATATTATTATTCAGCTTATATGAAGGAAGTATTTGACATG 65
QY      173 SerArgGln 175
Db      64 GTGCAGCAG 56
RESULT 13
; Sequence 18, Application US/09049698
; Patent No. 6368792
; GENERAL INFORMATION:
; APPLICANT: BILLING-MEDEL, PATRICIA A.
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: HAYDEN, MARK
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: ROBERTS-RAPP, LISA
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STROUPE, STEPHEN D.
; TITLE OF INVENTION: REAGENTS AND METHODS FOR THE
; TITLE OF INVENTION: USEFUL FOR DETECTING DISEASES OF THE GASTROINTESTINAL
; NUMBER OF SEQUENCES: 51
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/049,698
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/828,856
; FILING DATE: 31-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6068.US.P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3043 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-049-698-16
Alignment Scores:
Pred. No.: 18 6 Length: 3043
Score: 75.50 Matches: 32
Percent Similarity: 43.09% Conservative: 21
Best Local Similarity: 26.02% Mismatches: 49
Query Match: 6.95% Indels: 21
DB: Gaps: 5
US-09-966-880A-8 (1-198) x US-09-049-698-16 (1-3043)
QY      58 GluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCys----- 75
Db      376 GAACCTGGTGGTGTATGGTTCATCTCTACCTGGAG-----TGTAGGTGG 332
QY      76 -----TyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHis 93
Db      331 TGCACACTATAACATCAGCATGTTATGGTTTTCANGCTTTTGGCCCTTTTGTACTGAGGATT 272
```

```
Alignment Scores:
Pred. No.: 19.8 Length: 3181
Score: 75.50 Matches: 32
Percent Similarity: 43.09% Conservative: 21
Best Local Similarity: 26.02% Mismatches: 49
Query Match: 6.95% Indels: 21
DB: 4 Gaps: 5

US-09-966-880A-8 (1-198) x US-09-049-698-18 (1-3181)
QY 58 GluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCys----- 75
Db 387 GAACTGCTGGTGTGGTTCATCTCTACCTGGAG-----TGTAGGTGG 343
QY 76 -----TyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHis 93
Db 342 TGCAACTATTAACATCAGCATGTTATGCTTTTCATGTTTGGCCTTTTGTACTGAGGATT 283
QY 94 ValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeu 113
Db 282 TTCTTCCCAATTCCTCAGG-----AATTAATATAGATACATTTTGGAAAAAATCT 232
QY 114 TyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGly 133
Db 231 TTTTCTGT-----GGCTTCAACACAGGTACGTAGAAGCTGT 196
QY 134 ValGluIleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheValGlu 153
Db 195 AGTCACCATATCTCTATTGTTCAATTATTTTTCATCTCTGGCACACTAGATCTAT 136
QY 154 AsnHisGluArgThrPheLysAla---TrpGluGlyLeuHisGluAsnSerValArgLeu 172
Db 135 ACAAATGACAATATCTTCAAGGCCATTATTATTACAGCTTAATGAAGCAAGTATTGACTG 76
QY 173 SerArgGln 175
Db 75 GTGCAGCAG 67

RESULT 14
US-08-200-807-1
; Sequence 1, Application US/08200807
; Patent No. 5573939
; GENERAL INFORMATION:
; APPLICANT: B vik, Claes Olof, Eriksson, Ulf
; TITLE OF INVENTION: Isolated Protein Receptors, Antibodies which
; bind Thereto, Nucleic Acid Sequence Coding
; Patent No. 5573939
; TITLE OF INVENTION: Therefor, And Uses Thereof
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: PC-DOS
; SOFTWARE: Wordperfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/200,807
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/883,539
; FILING DATE: 15-MAY-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hanson, No. 5573939man D.
; REGISTRATION NUMBER: 30,946
; REFERENCE/DOCKET NUMBER: LUD 280
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 688-9200
; TELEFAX: (212) 838-3884
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2629 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: no
; ANTI-SENSE: no
; US-08-200-807-1
Alignment Scores:
Pred. No.: 26.8 Length: 2629
Score: 73.50 Matches: 29
Percent Similarity: 37.93% Conservative: 26
Best Local Similarity: 20.00% Mismatches: 43
Query Match: 6.77% Indels: 47
DB: 1 Gaps: 8

US-09-966-880A-8 (1-198) x US-08-200-807-1 (1-2629)
QY 57 ValGluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyr 76
Db 782 GTCAAAATTAATCTGTTCAAGTTTCTTCTTCATGGAGTCTT----- 823
QY 77 ArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArg----- 92
Db 824 -----TGGGAGGCCAATTATCATGATTGTTTGAATCCAATGAACCATGGGGTT 874
QY 93 -----HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThr 110
Db 875 TGCCTTCATATTGCTGACAAAAAAGA----- 901
QY 111 AlaArgLeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHis 130
Db 902 ---AAAAAGTATATATCAATAATAACAGGACCTCTCTTTTAACCTC-----TTTCAT 952
QY 131 ArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyr----- 146
Db 953 CAC-----ATCAATACCTATGAAGACCATGATTTCTGATTGTGGAT 994
QY 147 -----CysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAla----- 161
Db 995 CTCTGTTGCTGGAAGGATTGAAATTTGTTTATATTATTATTATTAGCCAATTTACGT 1054
QY 162 -----TrpGluGlyLeuHisGluAsnSerValArgLeuSerArg---GlnLeuArgArg 178
Db 1055 GAGAACTGGGAAGAGGTGAAAAAATGCCAGAAAGGCTCTCAGCCTGAAGATTAGGAGA 1114
QY 179 IleLeuLeuProLeu 183
Db 1115 TAGCTACTTCTCTTG 1129

RESULT 15
US-08-488-305A-1
; Sequence 1, Application US/08488305A
; Patent No. 5679772
; GENERAL INFORMATION:
; APPLICANT: B vik, Claes Olof, Eriksson, Ulf; Peterson, Per A.
; TITLE OF INVENTION: Isolated Protein Receptors, Antibodies which
; bind Thereto, Nucleic Acid Sequence Coding
; Patent No. 5679772
; TITLE OF INVENTION: Therefor, And Uses Thereof
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Felfe & Lynch
; STREET: 805 Third Avenue
; CITY: New York City
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
```

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 kb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: PC-DOS  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/488,305A  
FILING DATE: 7-JUNE-1995  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Kohli, Vineet  
REGISTRATION NUMBER: 37,003  
REFERENCE/DOCKET NUMBER: LUD 5280.3  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 688-9200  
TELEFAX: (212) 838-3884  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2629 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
HYPOTHETICAL: no  
ANTI-SENSE: no  
US-08-488-305A-1

Alignment Scores:  
Pred. No.: 26.8 Length: 2629  
Score: 73.50 Matches: 29  
Percent Similarity: 37.93% Conservative: 26  
Best Local Similarity: 20.00% Mismatches: 43  
Query Match: 6.77% Indels: 47  
DB: 1 Gaps: 8

US-09-966-880A-8 (1-198) x US-08-488-305A-1 (1-2629)

QY	57	ValGlulLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyr	76
Db	782	GTCAAAATTAATCTGTTCAAGTTCTTTCTCATGGAGTCTT-----	823
QY	77	ArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArg-----	92
Db	824	-----TGGGGAGCCAATTACATGGATTGTTTGAATCCAATGAACCATGGGGTT	874
QY	93	-----HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThr	110
Db	875	TGGCTTCATATTGCTGACAAAAAAGA-----	901
QY	111	AlaArgLeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHis	130
Db	902	---AAAAAGTATATCAATAATAAATACAGGACCTCTCCTTTTAACCTC-----TTTCAT	952
QY	131	ArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyr-----	146
Db	953	CAC-----ATCAATACCTATGAGACCATGAGTTTCTGATTGGGAT	994
QY	147	-----CysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAla-----	161
Db	995	CTCTGTTGCTGGAAAGATTGGAATTGTTTATTAATTATTTATTTAGCCAAATTACGT	1054
QY	162	-----TrpGluGlyLeuHisGluAsnSerValArgLeuSerArg---GlnLeuArgArg	178
Db	1055	GAGAACTGGGAAGAGGTGAAAAAATGCCAGAAAGGCTCTCTCAGCCTGAAGTTAGGAGA	1114
QY	179	IleLeuLeuProLeu	183
Db	1115	TACGTACTTCCTTTG	1129

Search completed: June 19, 2003, 19:04:05  
Job time : 59 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus.p2n model

Run on: June 19, 2003, 17:54:11 ; Search time 156 Seconds  
(without alignments)  
1862.504 Million cell updates/sec

Title: US-09-966-880A-8

Perfect score: 1086

Sequence: 1 MDLLMNRKFLYQKNVRW.....ILLPLYVDLDFARTLGL 198

Scoring table: BLOSUM62

Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 1042519 seqs, 733713590 residues

Total number of hits satisfying chosen parameters: 2085038

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlh  
-O=/cgn2\_1/USPTO\_spool/US09966880/runat\_14062003\_175646\_16343/app\_query.fasta\_1.391  
-DB=PublishedApplications\_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-NAPLEN=200000000 -USER=US09966880@cgn\_1.1.57\_erunat\_14062003\_175646\_16343  
-ICPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1086	100.0	2818	9 US-09-966-880A-7	Sequence 7, Appli
2	1008	92.8	2440	9 US-09-966-880A-1	Sequence 1, Appli
3	644.5	59.3	6564	9 US-09-966-880A-10	Sequence 10, Appl
4	644.5	59.3	11204	9 US-09-966-880A-35	Sequence 35, Appl

5	505	46.5	271	9 US-09-966-880A-13	Sequence 13, Appl
6	390	35.9	1534	10 US-09-729-674-173	Sequence 173, App
7	369.5	34.0	987	10 US-09-925-300-699	Sequence 699, App
8	369.5	34.0	1143	9 US-10-098-841-268	Sequence 268, App
9	363.5	33.5	1348	10 US-09-880-107-3282	Sequence 3282, Ap
10	308.5	28.4	2151	10 US-09-822-830A-359	Sequence 359, App
11	260	23.9	148	9 US-09-966-880A-12	Sequence 12, Appl
12	257	23.7	371	9 US-09-918-995-30237	Sequence 30237, A
13	238.5	22.0	1404	9 US-10-198-846-13472	Sequence 13472, A
14	218.5	20.1	879	9 US-10-157-031-13	Sequence 13, Appl
15	212	19.5	116	9 US-09-966-880A-14	Sequence 14, Appl
16	206	19.0	390	9 US-09-796-692-9051	Sequence 9051, Ap
17	206	19.0	390	9 US-10-040-862-9051	Sequence 9051, Ap
18	175	16.1	556	9 US-09-796-692-5061	Sequence 5061, Ap
19	175	16.1	556	9 US-10-040-862-5061	Sequence 5061, Ap
20	166.5	15.3	476	9 US-09-918-995-32385	Sequence 32385, A
21	164	15.1	553	9 US-09-796-692-4504	Sequence 4504, Ap
22	164	15.1	553	9 US-10-040-862-4504	Sequence 4504, Ap
23	162.5	15.0	572	10 US-09-864-761-9416	Sequence 9416, Ap
24	157.5	14.5	328	9 US-09-796-692-5737	Sequence 5737, Ap
25	157.5	14.5	328	9 US-10-040-862-5737	Sequence 5737, Ap
26	122	11.2	505	9 US-09-796-692-4443	Sequence 4443, Ap
27	122	11.2	505	9 US-10-040-862-4443	Sequence 4443, Ap
28	114	10.5	261	10 US-09-777-564-763	Sequence 763, App
29	114	10.5	261	10 US-10-015-219-763	Sequence 763, App
30	112.5	10.4	466	9 US-09-796-692-4007	Sequence 4007, Ap
31	112.5	10.4	466	9 US-10-040-862-4007	Sequence 4007, Ap
32	101	9.3	559	10 US-09-864-761-9458	Sequence 9458, Ap
33	99.5	9.2	473	9 US-10-046-935-1443	Sequence 1443, Ap
34	99.5	9.2	473	9 US-09-878-178-1443	Sequence 1443, Ap
35	99.5	9.2	473	9 US-10-146-502-1443	Sequence 1443, Ap
36	97.5	9.0	254	10 US-09-867-701-3576	Sequence 3576, Ap
37	88	8.1	2172	9 US-09-966-880A-15	Sequence 15, Appl
38	84	7.7	154	10 US-09-864-761-25835	Sequence 25835, A
39	84	7.7	997	10 US-09-864-761-9278	Sequence 9278, Ap
40	82.5	7.6	415	10 US-09-764-877-3699	Sequence 3699, Ap
41	81.5	7.5	14708	9 US-10-239-676-221	Sequence 221, App
42	80	7.4	555	10 US-09-864-761-9282	Sequence 9282, Ap
43	78	7.2	342	10 US-09-783-590-5098	Sequence 5098, Ap
44	77.5	7.1	1437	9 US-10-166-087-49	Sequence 49, Appl
45	77.5	7.1	1992	9 US-10-080-960-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1  
US-09-966-880A-7  
; Sequence 7, Application US/09966880A  
; Patent No. US2002016473A1  
; GENERAL INFORMATION:  
; APPLICANT: Honjo, Tasuku  
; APPLICANT: Muramatsu, Masamichi  
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE  
; FILE REFERENCE: 06501-088001  
; CURRENT APPLICATION NUMBER: US/09/966,880A  
; CURRENT FILING DATE: 2001-09-28  
; PRIOR APPLICATION NUMBER: PCT/JP00/01918  
; PRIOR FILING DATE: 2000-03-28  
; PRIOR APPLICATION NUMBER: JP 11-371382  
; PRIOR FILING DATE: 1999-12-27  
; PRIOR APPLICATION NUMBER: JP 11-178999  
; PRIOR FILING DATE: 1999-06-24  
; PRIOR APPLICATION NUMBER: JP 11-87192  
; PRIOR FILING DATE: 1999-03-29  
; NUMBER OF SEQ ID NOS: 36  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 7  
; LENGTH: 2818  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (80)...(673)





Qy	83	SerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnPro	102
Db	2682	TCCTGGAGCCCTGCTACGACTGTGCCGACATGTGCCGACITTCCTCGCAGGGAACCCC	2741
Qy	103	AsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGlu	122
Db	2742	AACCTCAGTCTGAGGATCTTCACCGCGCCTCTACTCTGTGAGGACCCCAAGGCTGAG	2801
Qy	123	ProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLys	142
Db	2802	CCCGAGGGGCTGGCGCGCTGCACCGCGCGGGTGCAATAGCCATCATGACCTTCAA	2860
Qy	142	-----	142
Db	2861	AGGTGCCAAAGGGCCTTCGCGCAGGGCGCAGTGCAGCAGCCGCAATTCGGGATTCGATG	2920
		-----	142
		-----	
Db	2921	CGGAATGAATGAGTTAGTGGGAAGCTCGAGGGGAAGAAGTGGCGGGGATTCGTGTTCA	2980
Qy	142	-----	142
		-----	
		-----	
Db	2981	CCTCTGAGCCGAAATTAAGATTAGNAGCAGAGAAAAGAGTGAATGGCTCAGAGACAAG	3040
Qy	142	-----	142
		-----	
		-----	
Db	3041	GCCTCGAGGAAATGAGAAAATGGGGCCAGGGTTGCTTCTTCCCTCGATTTGGAAACCTG	3100
Qy	143	-----	144
		-----	AspTyr
Db	3101	AACTGCTTCTACCCCCATATCCCGCGCTTTTTCCTTTTTCCTTTTTCGAGGATTTAT	3160
Qy	145	PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly	164
Db	3161	TTTTTACTGCTGGAATACTTTTGTAGAAAACCCAGAAAGACTTTTCAAAGCCTCGGGAAGGG	3220
Qy	165	LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeu	181
Db	3221	CTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTG	3271

```

RESULT 4
US-09-966-880A-35
; Sequence 35, Application US/0966880A
; Patent No. US2002016473A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966.880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 35
; LENGTH: 11204
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-35

```

Alignment Scores:	
Pred. No.:	7.11e-79
Score:	644.50
Length:	11204
Percent Similarity:	177
Best Local Similarity:	24.02%
Conservative:	0
Mismatches:	2

Query Match: DB:	59.35% 9	Indels: Gaps:	560 2
US-09-966-880A-8 (1-198) x US-09-966-880A-35 (1-11204)			
QY	3	SerLeuLeuMetAsnArgAtgLysPheLeuTyrcInPheLysAsnValArGTpAlaLys	22
Db	6278	AGCCTCTTGATGAACCGGAGAAATTTCTTTACCAATTCAAAAATGTCCTGGCTAAG	6337
QY	23	GLYArgATrGluThrTyrLeuCYsTyTyrValValLysArgArgAspSerAlaThrSerPhe	42
Db	6338	GGTCGGCGTGAGACCTACCTGTGCTACGTACTGNAAGAGCGGTGACAGTGTACATCCTTT	6397
QY	43	SerLeuAspPheGlyTyrLeuArgAsnLys	52
Db	6398	TCACGTGGACTTTGGTTATCTTCGCAATAA--GGTATCAATTAAGTCAGCTTTGCAAGCAG	6456
QY	52	-----	52
Db	6457	TTTAAATGGTCAACTGTGAGTCTTTTAGAGCCACCTGCTGATGTAATTACTTCCATCCTT	6516
QY	52	-----	52
Db	6517	TTTTGGCAATTTGTGCTCTATCACATTCCTCAAAATCCTTTTTTATTTCTTTTCCATG	6576
QY	52	-----	52
Db	6577	TCCATGCACCCATATTAGACATGCCCAAAATATGTGATTAAATTCCTCCCAGTAATGC	6636
QY	52	-----	52
Db	6637	TGGGCACCCTAATACCACTCCTTCTTTCAGTGCCAGAACAACTGCTCCCAAACTGTTA	6696
QY	52	-----	52
Db	6697	CCAGCTTTCCTCAGCATCTGAATTCCTTTGCCTTTTGAGATTAATTAAAGCTAAAGCATTTTATA	6756
QY	52	-----	52
Db	6757	TGGGAGAATAATTATCAGCTTGTCCAGCAAAAAATTTTAAATGTGAAAAACAATTTGTGC	6816
QY	52	-----	52
Db	6817	TTAAGCATTTTGGAAATTAAGGAGAGAGAAATTTGGGAAAAAATTAACGGTGGTTCAATT	6876
QY	52	-----	52
Db	6877	CTGTTTCCAAATGATTTCTTTTCCCTCTACTACATGGGTGCTAGGCCAGTGAATACA	6936
QY	52	-----	52
Db	6937	TTCAACATGTGTATCCCCAGAAAACTCAGAGAAGCCTCGGCTGATGATTAATTAAATTGA	6996
QY	52	-----	52
Db	6997	TCTTTCGGCTACCGGAGAGAAATTACATTTTCCAAGACACTTCTTCAACAAAATCCAGATGG	7056
QY	52	-----	52
Db	7057	GTTTACATAAACTCTGCCCATGGGTATCTCCTCTCTCTAACACGCTGTGAGCTCTGGG	7116
QY	52	-----	52
Db	7117	CTTGGTGGAAATCTCAGGGAAGCATCCGTGGGTGGAAGTCATCGCTCGCTGTGTATT	7176
QY	52	-----	52
Db	7177	GATGGTTATATTACCATGCAATTTTCTTTGGCCTACATTTGTATTGAAATACATCCCAATCT	7236
QY	52	-----	52
Db	7237	CCTTCCTATTTCGGTGACATGACACATTTCTATTTCAGAAGGCTTTTGATTTTATCAAGCACT	7296

```
QY 52 ----- 52
Db 7297 TTCAATTACTTCTCATGGCAGTGCCTATTACTTCTCTTACAATAACCCATCTCTGCTTT 7356
QY 52 ----- 52
Db 7357 ACCAAAATCTATTTCCTCCCTTTTCAGATCCCTCCAAATGGTCTCTCATAACTGCTCTGCT 7416
QY 52 ----- 52
Db 7417 CCACCTAGTGGTCCAGGTATATTCCACAATGTTACATCAACAGGCACCTTCTAGCCATT 7476
QY 52 ----- 52
Db 7477 TCCTTCTCAAAAGGTGCAAAAAGCAACTTCATAAACAACAATAATAATCTTCGGTGAGTA 7536
QY 52 ----- 52
Db 7537 GTGTGATGCTGCTTCTCCAACTCAGCGCACTTCGTCTTCCTCATTTCCAAAAACCCA 7596
QY 52 ----- 52
Db 7597 TAGCCTTCTTCACTCTGCAGCACTAGTCTGCCAAGGTTTCAGCTCTACCTACTGGTGT 7656
QY 52 ----- 52
Db 7657 GCTCTTTTGAGCAAGTTGCTTAGCCTCTCTGTAACACAGGACAATAGCTGCAAGCATCC 7716
QY 52 ----- 52
Db 7717 CCAAGATCATTGCAGGAGACAATGACTAAGGCTACCAGCGCCGAATAAAAGTCAGTGA 7776
QY 53 ----- AsnGlyCysHisValGluLeuLeuPheLeu 62
Db 7777 ATTTTAGCGTGTCTCTCTGTCTCTCCAGAACGGCTGCCAGCTGGAATGTCTCTCCTC 7836
QY 63 ArgTyrIleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThr 82
Db 7837 CGCTACATCTCGGACTGGGACCTAGACCTCGCCGCTGCTACCGCTCACCTGGTTCACC 7896
QY 83 SerTrpSerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnPro 102
Db 7897 TCCTGGAGCCCTCTAGCACTGTGCCGACATGTGGCGGACTTTCTGCGAGGGAACCC 7956
QY 103 AsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGlu 122
Db 7957 AACCTCAGTCTGAGGATCTTACCGCGGCTTACTTCTGTGAGGACCGCAAGGCTGAG 8016
QY 123 ProGluGlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLys 142
Db 8017 CCGAGGGGCTGCGCGGCTGCACCGCGCGGGTGCAAAATAGCCATCATGACCTTCAA- 8075
QY 142 ----- 142
Db 8076 AGGTCGGAAGGGCCCTTCGCGCAGGCGCAGTGCAGACGCCGCGCATTCGGGATTCGGATG 8135
QY 142 ----- 142
Db 8136 CGGAATGAATGATAGTGGGAAGCTCGAGGGGAAGAGTGGCGGGGATTCGTGTTCA 8195
QY 142 ----- 142
Db 8196 CCTCTGAGCCGAAATTAAGATTAGAACGACAGAGAAAAGTGAATGGCTCAGACACAAG 8255
QY 142 ----- 142
Db 8256 GCCCGGAGGAATAGAAAATGGGGCCAGGGTGTCTTCTCCCTCGATTTGGAACCTG 8315
QY 143 ----- AspTyr 144
Db 8316 AACTGTCTTCTACCCCATATCCCGCCTTTTTTTCCTTTTTTTTTTTTGAAGATTAT 8375
QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164
```

```
Db 8376 TTTTACTGCTGAATACTTTTGTAGAAAAACCAAGAAAGAACTTTTCAAAGCCTGGAGGG 8435
QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeuLeu 181
Db 8436 CTGCATGAAATTCAGTTCGTCTCTCCAGACAGCTTCGGCGCATCCTTTTG 8486

RESULT 5
US-09-966-880A-13
; Sequence 13, Application US/09966880A
; Patent No. US2002016473A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; PRIOR FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13
; LENGTH: 271
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-13

Alignment Scores:
Pred. No.: 9,39e-62 Length: 271
Score: 505.00 Matches: 90
Percent Similarity: 100.00% Conservatives: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 46.50% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-13 (1-271)
QY 53 AsnGlyCysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspPro 72
Db 1 AACGGCTGCCAGCTGGAATGTCTCTCCGCTACATCTCGGACTGGGACCTAGACCCT 60
QY 73 GlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArg 92
Db 61 GGCGGCTGTACCGGCTCACCTGGTTACCTCTCTGGAGCCCTGTCTACGACTGTGCCGA 120
QY 93 HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArg 112
Db 121 CATGTGGCGGACTTTCTGCGAGGGAACCCCAACCTCAGTCTGAGGATCTTCACCGCGCG 180
QY 113 LeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAla 132
Db 181 CTCTACTTCTGTGAGGACCGCAAGGCTGAGCCCGAGGGCTGCGGGCGGCTGCACCGGCC 240
QY 133 GlyValGlnIleAlaIleMetThrPheLys 142
Db 241 GGGGTGCAATAGCCATCATGACCTTCAAA 270

RESULT 6
US-09-729-674-173
; Sequence 173, Application US/09729674
; Patent No. US2001003935A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
```

```

; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fecthel, Kim
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6055-64x
; CURRENT APPLICATION NUMBER: US/09/729,674
; CURRENT FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: 09/539,330
; PRIOR FILING DATE: 2000-03-30
; NUMBER OF SEQ ID NOS: 283
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 173
; LENGTH: 1534
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-729-674-173

Alignment Scores:
Pred. No.: 2,06e-44 Length: 1534
Score: 390.00 Matches: 83
Percent Similarity: 61.62% Conservative: 31
Best Local Similarity: 44.86% Mismatches: 59
Query Match: 35.91% Indels: 12
DB: 10 Gaps: 3

US-09-966-880A-8 (1-198) x US-09-729-674-173 (1-1534)
QY 6 MetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgArg 25
Db ATGGATCCACCCACCTTCACTTCAACTTAAACATGAACCTTGGGTCAGAGCGGCAT 698
QY 26 GluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSerLeuAsp 45
Db GAGACTTACCTGTGTATGAGGTGGAGCGCATGCACATGACACCTGGGTCCTGCGAAC 758
QY 46 -----PheGlyTyrLeuArgAsnLys-----AsnGlyCys 55
Db CAGCGCAGGGCTTCTATGCAACACGAGGCTCCACATAAACACGGTTTCTTGAAGGCCGC 818
QY 56 HisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArgCys 75
Db CATGACAGAGCTGTGCTCTCTGGACGTGATTCCTTTTGGAAAGCTGGACCTGGACCGAG 878
QY 76 TyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaArgHisValAla 95
Db TACAGGGTTACCTGTTCACTTCTGGAGCCCTGCTTCACTGTGTCAGGAAATGGCT 938
QY 96 AspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArgLeuTyrPhe 115
Db AAATTCATTTCAAAAAACAAACAGCGTGCGCTGTGCATCTTCACTGCCCGCATCAT--- 995
QY 116 CysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAlaGlyValGln 135
Db ---GATGATCAAGGAAGATGTCTAGGAGGGGCTGCGCACCTTCTGGCCGAGCTGGGCCAAA 1052
QY 136 IleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheValGluAsnHis 155
Db ATTTCAATATGACATACACTGAAATTAAGCACTGCTGGGACACCTTTGTGGACCAACG 1112
QY 156 GluArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSerValArgLeuSerArgGln 175
Db GGATGTCCTTCCAGCCCTGGGATGGACTAGATGACACAGCAAGACCTGATGGGAGG 1172
QY 176 LeuArgArgIleLeu 180
Db CTGCGGGCCATTCTC 1187
```

## RESULT 7

```

US-09-925-300-699
; Sequence 699, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben,
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1890
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 699
; LENGTH: 987
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-925-300-699
```

```

Alignment Scores:
Pred. No.: 7.95e-42 Length: 987
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 34.02% Indels: 9
DB: 10 Gaps: 4
```

US-09-966-880A-8 (1-198) x US-09-925-300-699 (1-987)

```

QY 11 PheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgGluThrTyrLeuCys 30
Db TTTCTACTTCCAAATTTAAAAACCTATATGGGAAGCCACGATCGGAAGCAAACTTTGGTGTGC 206
QY 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48
Db TTTACCGTGGGAAGGTATAAAGCGCGCTCAGTTGTCTCTCTGGAAAGC-----GGCGTC 260
QY 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuLeuPheLeuArgTyr 64
Db TTTCCGAAACACGAGTGGATTCTCGAGCCCATTTCTCATGAGAAAGGTGCTTCTCTCTTGG 320
QY 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84
Db TTTCTGCGACGACATATCTGTCTCTTAACACAAAGTACCAGGTACCTGCTACACATCTTGG 380
QY 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104
Db ACCCTTGGCCAGACTGTGACGGGAGGTGGCGGAGTTCTCTGGCCAGGCACGACGAG 440
QY 105 SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu 124
Db AATCTCACCATCTTCCACCGCCGCTCTACTACTTCCAG---TATCCATGTTTACCAGGAG 497
QY 125 GlyLeuArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyr 144
Db GGGCTCCGCGAGCTGTAGTCAAGGAAGGGTTCGCTGGAGATCATGACTATCAAGATTTT 557
QY 145 PheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGly 164
Db AAATATGTTGGGAAAACTTTGTGTACATATGATATGAGCCCATTTCAAGCCTTGGAGGGA 617
QY 165 LeuHisGluAsnSerValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db TTTAAAAACAACACTTTTCGACTTCTGAAAGAGAGGCTACGGGAGAGTCTC 665

RESULT 8
US-10-098-841-268
; Sequence 268, Application US/10098841
```

```

; Publication No. US20020197679A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Xu, Chongjun
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhao, Qing A.
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wehrman, Tom
; APPLICANT: Zhang, Jie
; APPLICANT: Qian, Xiaohong B.
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: NO. US20020197679A1el Nucleic Acids and
; FILE REFERENCE: Polypeptides
; CURRENT APPLICATION NUMBER: US/10/098,841
; CURRENT FILING DATE: 2002-03-13
; PRIOR APPLICATION NUMBER: 09/598,042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 331
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 268
; LENGTH: 1143
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (128)..(700)
US-10-098-841-268

Alignment Scores:
Pred. No.: 9,98e-42 Length: 1143
Score: 369.50 Matches: 79
Percent Similarity: 58.52% Conservative: 24
Best Local Similarity: 44.89% Mismatches: 64
Query Match: 34.02% Indels: 9
DB: 9 Gaps: 4

US-09-966-880A-8 (1-198) x US-10-098-841-268 (1-1143)

Qy 11 PheLeuTyrGlnPheLeuAsnValArgTrpAlaLysGlyArgGluThrTyrLeuCys 30
Db 176 TTCACCTCCAACTTAAACCACTATGGGAAGCCCAACGATCGGAACGAACTGGCTGTGC 235
Qy 31 TyrValValLys-----ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyr 48
Db 236 TTCACCGTGGAGGTATAAAGCGCGCTCAGTTGCTCTCTCGGAAGACG-----GGCGTC 289
Qy 49 LeuArgAsn-----LysAsnGlyCysHisValGluLeuLeuPheLeuArgTyr 64
Db 290 TTCGGAACACAGGTGGATCTCGAGACCCATTGTCATGCAGAAAGGTCTCTCTCTTGG 349
Qy 65 IleSerAspTrpAspLeuAspProGlyArgCysTyrArgValThrTrpPheThrSerTrp 84
Db 350 TTCCTGCGAGCAGACTCTCTCTTAACAAAGTACCAGGTACCTGTGTACACATCTTGG 409
Qy 85 SerProCysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeu 104
Db 410 AGCCCTGCCAGACTGTGCAGGGGAGGTGCCGAGTTCCTGGCCAGCCAGCAACGTG 469
Qy 105 SerLeuArgIlePheThrAlaArgLeuTyrPheCysGluAspArgLysAlaGluProGlu 124
Db 470 AATCTCACCATCTTCACCGCGCGCTCTACTACTTCCAG---TATCCATGTTACCAGGAG 526

; Publication No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3282
; LENGTH: 1348
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U03891
; NAME/KEY: unsure
; LOCATION: (1)..(1348)
; OTHER INFORMATION: n = a o r c o r g o t
US-09-880-107-3282

Alignment Scores:
Pred. No.: 9e-41 Length: 1348
Score: 363.50 Matches: 83
Percent Similarity: 58.12% Conservative: 28
Best Local Similarity: 43.46% Mismatches: 57
Query Match: 33.47% Indels: 23
DB: 10 Gaps: 7

US-09-966-880A-8 (1-198) x US-09-880-107-3282 (1-1348)

Qy 5 LeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArg 24
Db 109 TTGATGGATCCACATATTCACCTCCCACTTTAACAAT-----GGCATTTGAAGG 159
Qy 25 ArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSerLeu 44
Db 160 CATAAGACCTACTCTGTCTACGAAGTGGAGCGCTGGACAATGGCACCTCGGTCAAGATG 219
Qy 45 Asp-----PheGlyTyrLeuArgAsn-----LysAsn-----Gly 54
Db 220 GACCAAGCACAGGGCTTTCTACACACACAGGCTAAGATCTCTCTGTGCTGTTCAGGC 279
Qy 55 CysHisValGluLeuLeuPheLeuArgTyrIleSerAspTrpAspLeuAspProGlyArg 74
Db 280 CGCCATCGGAGCTCGGCTTCTTGGACCTGGTTCTCTTTCAGTGTGGACCGCCAG 339
Qy 75 CysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAsp-----CysAlaArg 92
Db 340 ATCTACAGGGTCACCTGGTTTCATCTCTCTGGAGCCCTGCTCTCTCTCTCTCTCTCTG 399
```

```
QY 93 HisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaArg 112
Db 400 GAAGTGGCGTTCCTTCAGGAGAACACACACGTCGAGATCTTCGTCGTCGCCGC 459
QY 113 LeuTyrPheCysGluAspArgLysAlaGluPro-----GluGlyLeuArgArgLeu 129
Db 460 ATCTAT-----GATTACGACCCCTATATAAGGAGCGCGTCGAATGCTG 504
QY 130 HisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyrCysTyrAsn 149
Db 505 CGGGATGCTGGGCCCCAAGTCTCCATCATGACCTACGATGAATTAAGCACTGCTGGGAC 564
QY 150 ThrPheValGluAsnHisGluArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSer 169
Db 565 ACCTTTGTGGACCACACAGGATGTCCTCTCCAGCCCTGGGATGAGCTAGATGAGCACAGC 624
QY 170 ValArgLeuSerArgGlnLeuArgArgIleLeu 180
Db 625 CAAGCCCTGAGTGGGAGGCTGCGGGCCATTCTC 657

RESULT 10
; Sequence 359, Application US/09822830A
; Patent No. US20020142952A1
; GENERAL INFORMATION:
; APPLICANT: Genetics Institute, Inc.
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fecthel, Kim
; APPLICANT: Agostino, Michael J.
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6402
; CURRENT APPLICATION NUMBER: US/09/822,830A
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,604
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 631
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 359
; LENGTH: 2151
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-822-830A-359

Alignment Scores:
Pred. No.: 1.01e-32 Length: 2151
Score: 308.50 Matches: 81
Percent Similarity: 58.20% Conservative: 29
Best Local Similarity: 42.86% Mismatches: 60
Query Match: 28.41% Indels: 22
DB: 10 Gaps: 4

US-09-966-880A-8 (1-198) x US-09-822-830A-359 (1-2151)
QY 6 MetaAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGlyArgArg 25
Db 1278 ATGGATCCACCACCATTCACCTTTCAACTTTAAACAATGAACCTTGGGTGACAGA-CGGCAT 1336
QY 26 GluThrTyrLeuCystYrValValLysArg-----AspPheGlyTyrLeuArgAsnLy 37
Db 1337 GAGACTTACCTGTGTATGAGGTGGAGCGCATGCAACATGACACCTGGGTCTTGTCTGAA 1396
QY 38 SerAla-ThrSerPheSerLeu-----AspPheGlyTyrLeuArgAsnLy 52
Db 1397 CCAGCGCAGGGGCTTTCTTTATGACACCGAGGTCCACATAAACACACGCTTCTCTT----- 1448
QY 52 sAsnGlyCysHisValGluLeuPheLeuArgTyrIleSerAspTrpAspLeuAspPr 72
Db 1449 -GAAGGCCGCGATGCAGAGCTGTGCTCTCTGGAGCTGATCCCTTTTGGAGCTGGACCT 1507
```

```
QY 72 oGlyArgCysTyrArgValThrTrpPheThrSerTrpSerProCysTyrAspCysAlaAr 92
Db 1508 GGACCAGGACTACAGGTTACCTGTTACCTCTGGAGCCCTGCTTCAGCTGTGCCCA 1567
QY 92 gHisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeuArgIlePheThrAlaAr 112
Db 1568 GGAATGGCTAAATTCATTCTTCAAAAACAAACACGTCGAGCTGTGCATCTTTCACCTGCCGC 1627
QY 112 qLeuTyrPheCysGluAspArgLysAlaGluProGluGlyLeuArgArgLeuHisArgAl 132
Db 1628 CATCTAT-----GATGATCAAGGAAGATGTCAGAGGGGCTGGCCACCTTGGCCGAGGC 1681
QY 132 aGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyrCysTrpAsnThrPheVa 152
Db 1682 TGGGGCCAAAATTCATTAATGACATACAGTGAATTTAAGCACTGCTGGAC-ACCTTTGG 1740
QY 152 lGluAsnHisGluArgThrPheLysAlaTrpGluGlyLeuHisGluAsnSerValArgLe 172
Db 1741 -GACCACGAGGATGTCCTCTCCAGCCCTGGGATGGACTAGATGAGCACACGCAAGACCT 1799
QY 172 uSerArgGlnLeuArgArgIleLeu 180
Db 1800 GAGTGGGAGGCTGCGGGCCATTCTC 1824

RESULT 11
US-09-966-880A-12
; Sequence 12, Application US/09966880A
; Patent No. US20020164743A1
; GENERAL INFORMATION:
; APPLICANT: Honjo, Tasuku
; APPLICANT: Muramatsu, Masamichi
; TITLE OF INVENTION: NOVEL CYTIDINE DEAMINASE
; FILE REFERENCE: 06501-088001
; CURRENT APPLICATION NUMBER: US/09/966,880A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: PCT/JP00/01918
; PRIOR FILING DATE: 2000-03-28
; PRIOR APPLICATION NUMBER: JP 11-371382
; PRIOR FILING DATE: 1999-12-27
; PRIOR APPLICATION NUMBER: JP 11-178999
; PRIOR FILING DATE: 1999-06-24
; PRIOR APPLICATION NUMBER: JP 11-87192
; PRIOR FILING DATE: 1999-03-29
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 148
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-966-880A-12

Alignment Scores:
Pred. No.: 1.07e-27 Length: 148
Score: 260.00 Matches: 49
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 23.94% Indels: 0
DB: 9 Gaps: 0

US-09-966-880A-8 (1-198) x US-09-966-880A-12 (1-148)
QY 4 LeuLeuMetAsnArgArgLysPheLeuTyrGlnPheLysAsnValArgTrpAlaLysGly 23
Db 2 CTCCTTGATGAACCGGAGGAAGTTCTTTACCAATTCAAAAATGTCCGCTGGGCTAAGGT 61
QY 24 ArgArgGluThrTyrLeuCysTyrValValLysArgArgAspSerAlaThrSerPheSer 43
Db 62 CGGCGTGAGACCTACCTGCTAGTACGTAGTGAAGAGCGGTGACGTGCTACATCTTTTCA 121
QY 44 LeuAspPheGlyTyrLeuArgAsnLys 52
Db 122 CTGGACTTTGGTTATCTTCGCAATAAG 148
```



```

US-10-157-031-13
; Sequence 13, Application US/10157031
; Publication No. US2003010890A1
; GENERAL INFORMATION:
; APPLICANT: Baranova, A. V.
; APPLICANT: Yankovsky, N. K.
; APPLICANT: Kozlov, A. P.
; APPLICANT: Lobashev, A. V.
; APPLICANT: Krukovskaya, L. L.
; TITLE OF INVENTION: In silico screen

```

```

; FILE REFERENCE: 2760-103
; CURRENT APPLICATION NUMBER: US/10/157,031
; CURRENT FILING DATE: 2002-05-30
; NUMBER OF SEQ ID NOS: 415
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 879
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-157-031-13

```

Alignment Scores:		
Pred. No.:	1.15e-20	879
Score:	218.50	44
Percent Similarity:	55.5%	Conservative: 31
Best Local Similarity:	34.9%	Mismatches: 42
Query Match:	20.1%	Indels: 9
DB:	9	Gaps: 4

US-09-966-880A-8 (1-198) x US-10-157-031-13 (1-879)

QY		35	ArgArgAspSerAlaThrSerPheSerLeuAspPheGlyTyrLeuArg-----	50
Db		119	CCTAAAGAGCGCTGTCTCTACCAATCAAGTGGGCATGAGCCGAATGTCGCGA	178
			:	
QY		51	-----AsnLysAsnGlyCys---HisValIcluleuLeuPheLeu---ArgTyrIleSer	66
Db		179	AGCTCAGCGCAAAAACACCACAATCACGCTGGAGTTAATTTTATAAAAAAATTACGTCA	238
			:    :    :    :	
QY		67	AspTrpAspLeuAspProGlyArgCystyrArgValThrTrpPheThrSerTrpSerPro	86
Db		239	GNAAGAGATTTTACCACCATCATGACGCTGCCATCACCTGGTTCGTCTGGAGTCCC	298
			:    :    :    :    :    :    :    :	
QY		87	CysTyrAspCysAlaArgHisValAlaAspPheLeuArgGlyAsnProAsnLeuSerLeu	106
Db		299	TGCTGGGAATGCTCCAGGCTATTAGAGAGTTTCTGAGTCGGCACCCCTGGTGAGCTCTA	358
			:    :    :    :    :    :    :    :	
QY		107	ArgIlePheThrAlaArgLeutyrrPheCysGluAspArgLysalaGluProGluGlyLeu	126
Db		359	GTGATCTACGTAGCTCGCGCTTTTTTGGCACATGATGCACAA---AATCGGCAAGGTCTC	415
			:    :    :    :    :    :    :	
QY		127	ArgArgLeuHisArgAlaGlyValGlnIleAlaIleMetThrPheLysAspTyrPheTyr	146
Db		416	AGGGACCTTGTTACAGTGGAGTAAGTAATTTACAGATTATGAGACATCAGAGTATTATCAC	475
QY		147	CysTrpAsnThrPheVal	152
Db		476	TGCTGGAGGAATTTTGTCT	493

RESULT 15

US-09-966-880A-14  
; Sequence 14, Application US/09966880A  
; Patent No. US20020164743A1  
; GENERAL INFORMATION:

```

: GENERAL INFORMATION:
: APPLICANT: Hon'jo, Tasuku
: APPLICANT: Muramatsu, Masamichi
: TITLE OF INVENTION: NOVEL CITIDINE DEAMINASE
: FILE REFERENCE: 06501-088001
: CURRENT APPLICATION NUMBER: US/09/966.880A
: CURRENT FILING DATE: 2001-09-28
: PRIOR APPLICATION NUMBER: PCT/JP00/01918
: PRIOR FILING DATE: 2000-03-28
: PRIOR APPLICATION NUMBER: JP 11-371382
: PRIOR FILING DATE: 1999-12-27
: PRIOR APPLICATION NUMBER: JP 11-178999
: PRIOR FILING DATE: 1999-06-24
: PRIOR APPLICATION NUMBER: JP 11-87192
: PRIOR FILING DATE: 1999-03-29

```

```

; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 14
; LENGTH: 116
; TYPE: DNA

```

; ORGANISM: Homo sapiens  
US-09-966-880A-14

Alignment Scores:	4.12e-21	Length:	116
Pred. No.:	212.00	Matches:	38
Score:	212.00	Conservative:	0
Percent Similarity:	100.00%	Mismatches:	0
Best Local Similarity:	100.00%	Indels:	0
Query Match:	19.52%	Gaps:	0
DB:	9		

US-09-966-880A-8 (1-198) x US-09-966-880A-14 (1-116)

144	TyrPheTyrCysTrpAsnThrPheValGluAsnHisGluArgThrPheLysAlaTrpGlu	163
Qy		
3	TATTTTCTACTGCTGGAATACTTTTGTAGAAACACAGAAAGAACTTTCAAGGCTGGAA	62
Db		
164	GlyLeuHisGluAsnSerValArgLeuSerArgGlnLeuArgGlleLeuLeu	181
Qy		
63	GGCGTCGATGAAATTCAGTTCGTCCTCCAGACAGCTTCGGGCATCCCTTTG	116
Db		

Search completed: June 19, 2003, 19:06:49  
Job time : 161 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 23:33:05 ; Search time 14079.9 Seconds  
(without alignments)  
11397.322 Million cell updates/sec

Title: US-09-966-880A-9  
Perfect score: 5514  
Sequence: 1 acagagaatacatgtgtcca.....tcaaaactcctgacotcagag 5514

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	5482	99.4	71132	9	AC092184	AC092184 Homo sapi
2	4922.8	89.3	11204	9	AB040430	AB040430 Homo sapi
3	251.8	4.6	155922	9	AL160291	AL160291 Human DNA
4	243.2	4.4	161776	9	AL133229	AL133229 Human DNA
5	242.2	4.4	184672	2	AF002508	AF002508 Homo sapi
c 6	241.8	4.4	143517	9	AC083783	AC083783 Homo sapi
c 7	241.8	4.4	198582	9	AC005291	AC005291 Homo sapi
c 8	241.4	4.4	31632	9	AC104520	AC104520 Homo sapi
c 9	241.4	4.4	41407	9	AC053467	AC053467 Homo sapi
c 10	241.2	4.4	171073	2	AL590609	AL590609 Homo sapi
c 11	241.2	4.4	270269	2	AL590728	AL590728 Homo sapi
12	240.4	4.4	155085	2	AC025620	AC025620 Homo sapi
13	240.4	4.4	186107	2	AC023132	AC023132 Homo sapi
14	240.4	4.4	208164	2	AC113194	AC113194 Homo sapi
15	239.4	4.3	155521	9	AC069335	AC069335 Homo sapi
16	239.2	4.3	224048	2	AC023786	AC023786 Homo sapi
c 17	239	4.3	84170	9	HS130N4	275887 Human DNA s
c 18	239	4.3	123501	9	AL390241	AL390241 Human DNA
c 19	239	4.3	145166	2	AC053473	AC053473 Homo sapi
c 20	239	4.3	164810	2	AC023411	AC023411 Homo sapi
21	239	4.3	182653	9	AL354861	AL354861 Human DNA
c 22	239	4.3	266079	2	AL391060	AL391060 Homo sapi
c 23	238.8	4.3	66000	9	AL353665	AL353665 Human DNA
c 24	238.6	4.3	172611	9	AC019129	AC019129 Homo sapi
c 25	238.4	4.3	153665	2	AL138818	AL138818 Homo sapi
26	236.6	4.3	192104	2	AC090229	AC090229 Homo sapi
27	236.6	4.3	214978	9	AC007383	AC007383 Homo sapi
c 28	236	4.3	167758	2	AC068746	AC068746 Homo sapi
c 29	236	4.3	186299	2	AC107202	AC107202 Homo sapi
c 30	236	4.3	219554	2	AC107016	AC107016 Homo sapi
31	235.8	4.3	106896	9	AC074375	AC074375 Homo sapi
32	235.4	4.3	133226	2	AL512364	AL512364 Homo sapi
33	234.6	4.3	95719	2	AC015694	AC015694 Homo sapi
34	234.6	4.3	180776	2	AC032043	AC032043 Homo sapi
35	234.2	4.2	84570	9	AL359552	AL359552 Human DNA
36	234.2	4.2	162964	9	AC011939	AC011939 Homo sapi
37	234	4.2	158395	9	AC009242	AC009242 Homo sapi
c 38	233.4	4.2	142281	9	HS289880	AJ289880 Homo sapi
39	233.4	4.2	176968	9	AC005669	AC005669 Homo sapi
c 40	233.4	4.2	209772	9	AC099782	AC099782 Homo sapi
41	233.2	4.2	85500	9	HS199H16	AL022320 Human DNA
42	233	4.2	158838	9	AC009696	AC009696 Homo sapi
43	232.6	4.2	176343	9	CNS01DX3	AL139021 Human chr
44	232.4	4.2	168297	2	AC011271	AC011271 Homo sapi
c 45	232.4	4.2	170517	9	AC010545	AC010545 Homo sapi

ALIGNMENTS

RESULT 1  
AC092184  
LOCUS AC092184 Homo sapiens 71132 bp DNA linear PRI 12-JUN-2002  
DEFINITION BAC Library 12 BAC RP11-438L7 (Roswell Park Cancer Institute Human  
ACCESSION AC092184 AC013443 complete sequence.  
VERSION AC092184.7 GI:21206067  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 71132)  
AUTHORS Muzny D.M., Adams C., Adio-Oduola B., Ali-osman F.R., Allen C.,  
Alsbrooks S.L., Amarantunge H.C., Are J.R., Ayele M., Banks T.,

Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carter, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Haylak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogues, M., Holloway, C., Hollins, B., Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Loulseged, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Marondel, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mathew, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S., Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rivers, M., Rojas, A., Rojubokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shooshtari, N., Sisson, I., Sodergren, E., Sonaik, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Kucherlapati, R., Weinstock, G. and Gibbs, R.

Direct Submission  
2 (bases 1 to 71132)  
Worley, K.C.  
Direct Submission  
Submitted (25-JUN-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 71132)  
Worley, K.C.  
Direct Submission  
Submitted (18-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
4 (bases 1 to 71132)  
Worley, K.C.  
Direct Submission  
Submitted (25-MAY-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
5 (bases 1 to 71132)  
Worley, K.C.  
Direct Submission  
Submitted (12-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 25, 2002 this sequence version replaced gi:20901754.  
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email [gc-help@bcm.tmc.edu](mailto:gc-help@bcm.tmc.edu)

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of

the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the features listing.

#### ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8068/quality.info/genbank.annotation.html>.

#### QUALSTAT-REPORT.

FEATURES	source	Location/Qualifiers
		1..71132
		/organism="Homo sapiens"
		/db_xref="taxon:9606"
		/chromosome="12"
		/clone="RP11-438L7"
		complement(1..1448)
		/notes="overlaps bases 1..1448 of clone AC092490"
		/function="clone overlap"
STS		187..408
		/standard_name="57233"
STS		439..560
		/standard_name="92005"
repeat_region		774..881
		/rpt_family="MIR"
repeat_region		903..1190
		/rpt_family="AluSq"
repeat_region		1191..1213
		/rpt_family="AT_rich"
STS		1591..1807
		/standard_name="6612"
STS		1744..1819
		/standard_name="8198"
repeat_region		1966..2264
		/rpt_family="AluSq"
repeat_region		3296..3328
		/rpt_family="(TTTC)n"
repeat_region		4097..4249
		/rpt_family="AluSq"
repeat_region		4360..4654
		/rpt_family="AluY"
repeat_region		complement(5140..5262)
		/rpt_family="FLAM_C"
repeat_region		5275..5304
		/rpt_family="AT_rich"
repeat_region		5305..5563
		/rpt_family="AluJo"
repeat_region		5564..5599
		/rpt_family="AT_rich"
repeat_region		5644..5733
		/rpt_family="MSTD"

[illegible]



Db	37619	AACGCTGTTTCTATAACAAAATTAGCCGGGCATGGTGCATGTGCCTGTGTGCCAGCT	37678
QY	3721	ACTAGGGGCTGAGCAGAGAAATCTTTGGAGCCAGGAGGTCAAGGTGCACTGAGCAG	3780
Db	37679	ACTAGGGGCTGAGCAGAGAAATCGTTGGAGCCAGGAGGTCAAGGCTGCACCTGAGCAG	37738
QY	3781	TGCTTGCCGCTGACCTGACCTCCAGCCTGGGTGACAGCAGCAGACCTTGGCTCAAAAAATAA	3840
Db	37739	TGCTTGCCGCTGACCTCCAGCCTGGGTGACAGCAGCAGACCTTGGCTCAAAAAATAA	37798
QY	3841	GAAGAAAAATTAATAATAAATGGAACCAACTACAAAGAGCTTGTGCTAGATGAGCTAC	3900
Db	37799	GAAGAAAAATTAATAATAAATGGAACCAACTACAAAGAGCTTGTGCTAGATGAGCTAC	37858
QY	3901	TTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGCACCTGCACATAC	3960
Db	37859	TTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGCACCTGCACATAC	37918
QY	3961	ATTATTAATAATCAATTTCTCAATGTATATCCACACAAAGACTGGTACGTGATGTTCAAT	4020
Db	37919	ATTATTAATAATCAATTTCTCAATGTATATCCACACAAAGACTGGTACGTGATGTTCAAT	37978
QY	4021	AGTACCTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTGAA	4080
Db	37979	AGTACCTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATCAACAAGTGAA	38038
QY	4081	CAATTAACAAAATGTGCTATATCATGCAATGGAATACCAACCTGCAGTACAAAGGAAG	4140
Db	38039	CAATTAACAAAATGTGCTATATCATGCAATGGAATACCAACCTGCAGTACAAAGGAAG	38094
QY	4141	AAGCTACTTGGGATGAATCCCAAGTATGACGCTTAATGAAAGACTGACACATGAAGG	4200
Db	38095	AAGCTACTTGGGATGAATCCCAAGTATGACGCTTAATGAAAGACTGACACATGAAGG	38154
QY	4201	AGGAGATAATGTATGCCATAGCAAAATCTAGAAAATGAAAGTAACCTATATGTTACAGAAA	4260
Db	38155	AGGAGATAATGTATGCCATAGCAAAATCTAGAAAATGAAAGTAACCTATATGTTACAGAAA	38214
QY	4261	GCAATCAGGCGAGCATAGAGGCTCACACCTGTAAATCCCAGCACTTTGAGAGGCCACGT	4320
Db	38215	GCAATCAGGCGAGCATAGAGGCTCACACCTGTAAATCCCAGCACTTTGAGAGGCCACGT	38274
QY	4321	GGGAAGATTGCTAGAACTCAGAGTTCAAGACCGCTGGGCAACACAGTGAACCTCCAT	4380
Db	38275	GGGAAGATTGCTAGAACTCAGAGTTCAAGACCGCTGGGCAACACAGTGAACCTCCAT	38334
QY	4381	TCTCCACAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTGTCTGCTGTGGGAGGGGAAG	4440
Db	38335	TCTCCACAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTGTCTGCTGTGGGAGGGGAAG	38394
QY	4441	GACTGCAAGAGGGAAGAGCTCTGGTGGGTGAGGGTGGTGAATTCAGGTTCTGTATCCT	4500
Db	38395	GACTGCAAGAGGGAAGAGCTCTGGTGGGTGAGGGTGGTGAATTCAGGTTCTGTATCCT	38454
QY	4501	GACTGTGCTAGCAGTTGGGCTGTTTACATCCAAAATATTCGTAGAATTTATGCATCTTA	4560
Db	38455	GACTGTGCTAGCAGTTGGGCTGTTTACATCCAAAATATTCGTAGAATTTATGCATCTTA	38514
QY	4561	AATGGGTGGAGTTTACTGTATGTAAATATACCTCAATGTAAAGAAAAAATATGTGTAAG	4620
Db	38515	AATGGGTGGAGTTTACTGTATGTAAATATACCTCAATGTAAAGAAAAAATATGTGTAAG	38574
QY	4621	AAAAGTTTCAATTTCTTTGCCAGCAACAGTTTATCAAAATCCTGAGCCCTTTACTTCGCA	4680
Db	38575	AAAAGTTTCAATTTCTTTGCCAGCAACAGTTTATCAAAATCCTGAGCCCTTTACTTCGCA	38634
QY	4681	AATTCCTGCACTTCTGCCCCGTACCATTAGTGAACACACTAGCTCCACAAATTTGGATA	4740
Db	38635	AATTCCTGCACTTCTGCCCCGTACCATTAGTGAACACACTAGCTCCACAAATTTGGATA	38694
QY	4741	AATGCATTTCTGGAAAAAGACTAGGACAAAAATCCAGGCATCACTTGTGCTTTTCATATCAA	4800
Db	38695	AATGCATTTCTGGAAAAAGACTAGGACAAAAATCCAGGCATCACTTGTGCTTTTCATATCAA	38754
QY	4801	CCACGCTGTACAGCTTGTGTGCTGTCTGAGCTGCAATGGGGACTCTTGATTTCTTTTAA	4860
Db	38755	CCATGCTGTACAGCTTGTGTGCTGTCTGAGCTGCAATGGGGACTCTTGATTTCTTTTAA	38814
QY	4861	GCAGAACTTGGGTTTACCAGAGTATTTCCACAAATGCTATTTCAAAATAGTCTGATGATG	4920
Db	38815	GCAGAACTTGGGTTTACCAGAGTATTTCCACAAATGCTATTTCAAAATAGTCTGATGATG	38874
QY	4921	CAAGACACTGTCTAGGAGCCAGAAAAAAGAGGAGGAGAAATCAATGATTTATGTTGGGA	4980
Db	38875	CAAGACACTGTCTAGGAGCCAGAAAAAAGAGGAGGAGAAATCAATGATTTATGTTGGGA	38934
QY	4981	ACAACATAGCAAGATATTTTAGATCATTTTGTACTAGTTTAAAAAGCAGCAGAGTACAAAAT	5040
Db	38935	ACAACATAGCAAGATATTTTAGATCATTTTGTACTAGTTTAAAAAGCAGCAGAGTACAAAAT	38994
QY	5041	CACACATGCAATCAGTATTAATCCAAATCATGTAATATGTGCTGTAGAAAGACTAGAGG	5100
Db	38995	CACACATGCAATCAGTATTAATCCAAATCATGTAATATGTGCTGTAGAAAGACTAGAGG	39054
QY	5101	AATAAACACAAGAACTTTAAACAGTCAATGCTCATTTAGACACTTAAGTCTAATATTATTATT	5160
Db	39055	AATAAACACAAGAACTTTAAACAGTCAATGCTCATTTAGACACTTAAGTCTAATATTATTATT	39114
QY	5161	AGACACTATGATATTTTGAGATTTAAAAAATCTTTAATATTTTAAAAATTTAGAGCTCTTCT	5220
Db	39115	AGACACTATGATATTTTGAGATTTAAAAAATCTTTAATATTTTAAAAATTTAGAGCTCTTCT	39174
QY	5221	ATTTTTCATAGTATTTCAAGTTTGACATGATCAAGTATTAATCTTCTTTTCTTTTCTTTT	5280
Db	39175	ATTTTTCATAGTATTTCAAGTTTGACATGATCAAGTATTAATCTTCTTTTCTTTTCTTTT	39234
QY	5281	TTTTTTTTTTTTTTGAGATGGAGTTTGGTCTTTGTTGGCCCATGCTGGAGTGGAAATGGCATG	5340
Db	39235	TTTTTTTTTTTTTTGAGATGGAGTTTGGTCTTTGTTGGCCCATGCTGGAGTGGAAATGGCATG	39294
QY	5341	AYCATAGTCTACTGCAACCTCCACCTCTGGGTTCAAGCAAAAGCTGTGCGCTCAGCCTCC	5400
Db	39295	ACCATAGTCTACTGCAACCTCCACCTCTGGGTTCAAGCAAAAGCTGTGCGCTCAGCCTCC	39354
QY	5401	CGGTTAGATGGATACAGGCGCCCAACACACACACCTGCTGCTCAATGTTGTTGTTTATGTA	5460
Db	39355	CGGTTAGATGGATACAGGCGCCCAACACACACACCTGCTGCTCAATGTTGTTGTTTATGTA	39414
QY	5461	GAGATGGGTTTTCACCATGTTGGCAGGCTGCTCAAACTCCTGACCTCAGAG 5514	
Db	39415	GAGATGGGTTTTCACCATGTTGGCAGGCTGCTCAAACTCCTGACCTCAGAG 39468	
RESULT 2			
AB040430			
LOCUS			
DEFINITION			
Homo sapiens AID gene for activation-induced cytidine deaminase, complete cds.			
ACCESSION			
AB040430			
VERSION			
AB040430.1 GI:9988407			
KEYWORDS			
AID; activation-induced cytidine deaminase.			
SOURCE			
Homo sapiens DNA.			
ORGANISM			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE			
1 (sites)			
Muto, T., Muramatsu, M., Taniwaki, M., Kinoshita, K. and Honjo, T.			
Isolation, tissue distribution, and chromosomal localization of the			
human activation-induced cytidine deaminase (AID) gene			
Genomics 66 (1), 85-88 (2000)			
JOURNAL			
MEDLINE			
20408890			
REFERENCE			
2 (sites)			
Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O.,			
Catalan, N., Forveille, M., Dufourcq-Lagelouse, R., Gennery, A.,			
Tescan, I., Ersoy, F., Kayserili, H., Ugazio, A.G., Brousse, N.,			
Muramatsu, M., Notarangelo, L.D., Kinoshita, K., Honjo, T., Fischer, A.			



and Durandy, A.  
 Activation-induced cytidine deaminase (AID) deficiency causes the  
 autosomal recessive form of the Hyper-IgM syndrome (HIGM2)  
 Cell 102 (5), 565-575 (2000)  
 20460541  
 3 (bases 1 to 11204)  
 Muto, T., Muramatsu, M., Tanikawa, M., Kinoshita, K. and Honjo, T.  
 Direct Submission  
 Submitted (18-MAR-2000) Tasuku Honjo, Kyoto University, Department  
 of Medical Chemistry, Faculty of Medicine, Yoshida, Sakyo-ku,  
 Kyoto, Kyoto 606-8501, Japan (E-mail: honjo@med.kyoto-u.ac.jp,  
 Tel: 81-75-753-4371(ex. 4371), Fax: 81-75-753-4388)  
 Location/Qualifiers  
 1. .11204  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)  
 /gene="AID"  
 join(521..528,6280..6427,7807..8077,8371..8486,8956..9009)  
 /gene="AID"  
 /codon\_start=1  
 /product="activation-induced cytidine deaminase"  
 /protein\_id="BAB12720.1"  
 /db\_xref="GI:9988408"  
 /translation="MDSLLMNRKFLYQFKVNRWAKGRRETYLIVVKKRDSATSESL  
 DFGYLRNKGCHVEELFLRIISDWDLDPGRYRVTFWSVPCYDCARHVDLFRNP  
 NLSLRIFTARLYFEDRKAPEGLRLRHAGVQIAITPFKDYFCWNTPEVENHERTEK  
 AWEGLEHNSVRLSRLRILLPLYEVDDLRFALFTGL"  
 a 3305 a 2273 c 2373 g 3253 t

Query Match 89.3%; Score 4922.8; DB 9; Length 11204;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 4921; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 591 AGGTCAGAGACGCTGGGAATATGGGGAATAGAGGCTATCTGAGGCTCTTCAACAC 650  
 DB 1 AGGTCAGAGACGCTGGGAATATGGGGAATAGAGGCTATCTGAGGCTCTTCAACAC 60

QY 651 AATAACCCAAAGAGCTATTAAATGCTCTTTAAGGTATTACATAAATATTACTATTCTC 710  
 DB 61 AATAACCCAAAGAGCTATTAAATGCTCTTTAAGGTATTACATAAATATTACTATTCTC 120

QY 711 ATTGTCCTTTATTTGTTATCATGATTAATTAAGTGTCTACTGCTTACTGCTCC 770  
 DB 121 ATTGTCCTTTATTTGTTATCATGATTAATTAAGTGTCTACTGCTTACTGCTCC 180

QY 771 TGATCTTTGCTAGCTATGAGCATGGACTGGGCTTTTAGAGCAGCAGCCCAAGGAACC 830  
 DB 181 TGATCTTTGCTAGCTATGAGCATGGACTGGGCTTTTAGAGCAGCAGCCCAAGGAACC 240

QY 831 TAAACATTAAAGCAGAGCTGCCCTCAATGGTTTAACTGTTGACTCTGCCTATGACAGC 890  
 DB 241 TAAACATTAAAGCAGAGCTGCCCTCAATGGTTTAACTGTTGACTCTGCCTATGACAGC 300

QY 891 CCCACCCACCATCTTACTGGATCCAAATCAGGAGCAGCGGTGGGGTACCTGGTGG 950  
 DB 301 CCCACCCACCATCTTACTGGATCCAAATCAGGAGCAGCGGTGGGGTACCTGGTGG 360

QY 951 GGGTGATGCTGTCAGGGAGGAGCCCAAGGGCAGGCTCAAAATTTGAATGTAAGGGCC 1010  
 DB 361 GGGTGATGCTGTCAGGGAGGAGCCCAAGGGCAGGCTCAAAATTTGAATGTAAGGGCC 420

QY 1011 AATGCACTGTCAGACTGAGACAGAGAACCATCATTTAATTTGAAGTGAGATTTTCTGGCCT 1070  
 DB 421 AATGCACTGTCAGACTGAGACAGAGAACCATCATTTAATTTGAAGTGAGATTTTCTGGCCT 480

QY 1071 GAGACTTCAGGGAGGCAAGAGACACTCTGACACCACTATGACAGGTAAGAGGCGAG 1130  
 DB 481 GAGACTTCAGGGAGGCAAGAGACACTCTGACACCACTATGACAGGTAAGAGGCGAG 540

QY 1131 TCTTCTGCTGGTGAATGCACCTGGCCTTCTCTCAGACCAATCTGAGTAATGACACTGG 1190  
 DB 1131 TCTTCTGCTGGTGAATGCACCTGGCCTTCTCTCAGACCAATCTGAGTAATGACACTGG 1680

DB 541 TCCTTCGCGGGTGATGGCACTGGCCTTCCTCTCAGAGCAAAATCTGAGTAATGAGACTGG 600  
 QY 1191 TAGCTATCCCTTTCTCTCATGTAACTGCTCTGACTGATAAGACTCAGCTTTGATCAATATGCA 1250  
 DB 601 TAGCTATCCCTTTCTCTCATGTAACTGCTCTGACTGATAAGACTCAGCTTTGATCAATATGCA 660  
 QY 1251 TATATATTTTTTGTGATCTGCTCTCTTTCTTCTTATTCAGATCTTTATAGCGTGTGAGCCAA 1310  
 DB 661 TATATATTTTTTGTGATCTGCTCTCTTTCTTCTTATTCAGATCTTTATAGCGTGTGAGCCAA 720  
 QY 1311 TTCCTTCTGTTTCAGACTCTCTCTGATTTCCCTCTTTTTCATGTGGCAAAAGAGTAGTG 1370  
 DB 721 TTCCTTCTGTTTCAGACTCTCTCTGATTTCCCTCTTTTTCATGTGGCAAAAGAGTAGTG 780  
 QY 1371 CGTACAATGTACTGATTCGCTCGAGATTTGTACCATTGTTGAAACTAAATTTATGGTAAT 1430  
 DB 781 CGTACAATGTACTGATTCGCTCGAGATTTGTACCATTGTTGAAACTAAATTTATGGTAAT 840  
 QY 1431 AATATTAACTAGCAAAATCTTTAGAGACTCAAAATCATGAAAAGGTAAATAGCAGTACTGTA 1490  
 DB 841 AATATTAACTAGCAAAATCTTTAGAGACTCAAAATCATGAAAAGGTAAATAGCAGTACTGTA 900  
 QY 1491 CTAAAAACGCTAGTGTCTAAATTTTTCGTAATAATTTTCTGTAATATTTCAACAGTAAACAACT 1550  
 DB 901 CTAAAAACGCTAGTGTCTAAATTTTTCGTAATAATTTTCTGTAATATTTCAACAGTAAACAACT 960  
 QY 1551 TGAAGACACACTTTCCCTAGGAGCGCTTACTGAAATAATTTAGCTATAGTAAGAAAATTT 1610  
 DB 961 TGAAGACACACTTTCCCTAGGAGCGCTTACTGAAATAATTTAGCTATAGTAAGAAAATTT 1020  
 QY 1611 GTAATTTTGAATGCCAAGCATTTCTAAATTAATTTGCTTGAAGTCACTATGATGTGTC 1670  
 DB 1021 GTAATTTTGAATGCCAAGCATTTCTAAATTAATTTGCTTGAAGTCACTATGATGTGTC 1080  
 QY 1671 CATTTATAAGAGAGACAAATTCATTCAGCAAGTTTATTAAATTTAAAGGCCCAATTTGTTAG 1730  
 DB 1081 CATTTATAAGAGAGACAAATTCATTCAGCAAGTTTATTAAATTTAAAGGCCCAATTTGTTAG 1140  
 QY 1731 GCAGTTAATGGCACTTTTACTATTAACTAATCTTCCATTTGTTGACAGCTAGCTTAACT 1790  
 DB 1141 GCAGTTAATGGCACTTTTACTATTAACTAATCTTCCATTTGTTGACAGCTAGCTTAACT 1200  
 QY 1791 TACCTCTTAGGTGTGAATTTGGTTAAGTCCCTCATTAATGTCTTTATGTGCAAGTTTGTGAT 1850  
 DB 1201 TACCTCTTAGGTGTGAATTTGGTTAAGTCCCTCATTAATGTCTTTATGTGCAAGTTTGTGAT 1360  
 QY 1851 AGGTTATTCTCATAGAACTTATTCTATCTCATATTATGATTACTATGGATCTATGAGA 1910  
 DB 1261 AGGTTATTCTCATAGAACTTATTCTATCTCATATTATGATTACTATGGATCTATGAGA 1320  
 QY 1911 ATAACACCTTAATCTTTATCTTTTACCTCAATTTTAACTCTCTTTTAAAGAACTTACATTTAC 1970  
 DB 1321 ATAACACCTTAATCTTTATCTTTTACCTCAATTTTAACTCTCTTTTAAAGAACTTACATTTAC 1380  
 QY 1971 AGAATAAAGATTTTTTAAAAATATATTTTTTGTAGAGACAGGGTCTTAGCCAGCCGAG 2030  
 DB 1381 AGAATAAAGATTTTTTAAAAATATATTTTTTGTAGAGACAGGGTCTTAGCCAGCCGAG 1440  
 QY 2031 GCTGGTCTTAAGTCTGGCCCAAGCGATCTCTCTGGCCTCTTAAAGTCTGGA 2090  
 DB 1441 GCTGGTCTTAAGTCTGGCCCAAGCGATCTCTCTGGCCTCTTAAAGTCTGGA 1500  
 QY 2091 TTATAGACATGAGCCATCATCTCCATATACAGAAATAAAGATTTTTTAAATGGAGGATTTAA 2150  
 DB 1501 TTATAGACATGAGCCATCATCTCCATATACAGAAATAAAGATTTTTTAAATGGAGGATTTAA 1560  
 QY 2151 TGTCTTTCAGAAAAATTTTCTTGGGTGACAGCAATGTCAAAATGTCTCCTCAGTTTACACTG 2210  
 DB 1561 TGTCTTTCAGAAAAATTTTCTTGGGTGACAGCAATGTCAAAATGTCTCCTCAGTTTACACTG 1620  
 QY 2211 AGATTTTGAACAAAGCTCTGAGCTATAGCTCTTGTGAAGGGTCCCATTTGAATACTTCT 2270  
 DB 1621 AGATTTTGAACAAAGCTCTGAGCTATAGCTCTTGTGAAGGGTCCCATTTGAATACTTCT 1680



QY	2271	TCAAAGTAAATGGAAGCAAGGTAAATACAGCAGTTGAAATTCAGAGAAGACAGAAA	2330	3351	GTGTTGGATACAGTCATGAGCCACTGCATCTGGCTAGATCCCATTTAGATTAATAATAT	3410
Db	1681	TCAAAGTAAATGGAAGCAAGGTAAATACAGCAGTTGAAATTCAGAGAAGACAGAAA	1740	2761	GTGTTGGATACAGTCATGAGCCACTGCATCTGGCTAGATCCCATTTAGATTAATAATAT	2820
QY	2331	AGGAGAAAGATGAAATTCACAGACAGACAGAGGAAATATATATCATTAAGGAGGACAG	2390	3411	GCATTTTAAATTTTAAATAATATATGGCTAAATTTTACCTTATATGTTATGTTACGTTA	3470
Db	1741	AGGAGAAAGATGAAATTCACAGACAGACAGAGGAAATATATATCATTAAGGAGGACAG	1800	2821	GCATTTTAAATTTTAAATAATATATGGCTAAATTTTACCTTATATGTTATGTTACGTTA	2880
QY	2391	TATCTGTAGACCTCATTTAGTGTATGGCAAAATGACTTGTTCAGGATATATTTAACCCGCT	2450	3471	ATAAATCTAGTTTGTCTGCCTAAAGTTTAAAGTGCCTTCCAATAAGCTTCATGCTAGCTGAG	3530
Db	1801	TATCTGTAGACCTCATTTAGTGTATGGCAAAATGACTTGTTCAGGATATATTTAACCCGCT	1860	2881	ATAAATCTAGTTTGTCTGCCTAAAGTTTAAAGTGCCTTCCAATAAGCTTCATGCTAGCTGAG	2940
QY	2451	TGTTTCTGTTTGCACGGCTGGGATGCAGCTAGGTTCTGCCTCAGGGACACAGCTGT	2510	3531	GGGACACATTTAAAGTGAACACAGACAGCAGGCTGTGGTGCCTCAAGCTTAAATCCACG	3590
Db	1861	TGTTTCTGTTTGCACGGCTGGGATGCAGCTAGGTTCTGCCTCAGGGACACAGCTGT	1920	2941	GGGACACATTTAAAGTGAACACAGACAGCAGGCTGTGGTGCCTCAAGCTTAAATCCACG	3000
QY	2511	CCAGAGCAGCTGTGAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCCTTCCTACTCA	2570	3591	ACTCTGGAGGCTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCAAGCTGAGC	3650
Db	1921	CCAGAGCAGCTGTGAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCCCTTCCTACTCA	1980	3001	ACTCTGGAGGCTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCAAGCTGAGC	3060
QY	2571	GGACAGAAATGACGAGACAGGAGCTGGAACAGGCCCTTAACAGAGAAGGGAAGTAA	2630	3651	AACATGGCCAAACCCCTGTTCTATACAAAAATTTAGCCGGCATGGTGGCATGTGCCTGT	3710
Db	1981	GGACAGAAATGACGAGACAGGAGCTGGAACAGGCCCTTAACAGAGAAGGGAAGTAA	2040	3061	AACATGGCCAAACCCCTGTTCTATACAAAAATTTAGCCGGCATGGTGGCATGTGCCTGT	3120
QY	2631	TGGATCAACAAAGTTAACTAGCAGTCAAGATCAGCAATTCATTTCACTCTGACTGGTA	2690	3711	GGTCCAGCTACTAGGGGCTGAGCAGGAGAAATCTTTGGAGCCAGGAGGTCAAGGCTG	3770
Db	2041	TGGATCAACAAAGTTAACTAGCAGTCAAGATCAGCAATTCATTTCACTCTGACTGGTA	2100	3121	GGTCCAGCTACTAGGGGCTGAGCAGGAGAAATCTTTGGAGCCAGGAGGTCAAGGCTG	3180
QY	2691	ACATGTGACAGAAACAGTGTAGGCTTATGTTATTTTCATGTAGAGTAGGACCCAAAAATC	2750	3771	CACGTAGCAGTGTTCGCCACTGCACCTCCAGCCTGGGTGACAGACCAAGCTTGCCTC	3830
Db	2101	ACATGTGACAGAAACAGTGTAGGCTTATGTTATTTTCATGTAGAGTAGGACCCAAAAATC	2160	3181	CACGTAGCAGTGTTCGCCACTGCACCTCCAGCCTGGGTGACAGACCAAGCTTGCCTC	3240
QY	2751	CACCCAAAGTCCCTTATCTATGCCACATCTCTTATCTATCTATCTATCTATCTATCTATCT	2810	3831	AAAAAAATAAGAGAAAAATTTAAAAATAAATGGAACAACACTACAAGAGCTGTGTCTCTA	3890
Db	2161	CACCCAAAGTCCCTTATCTATGCCACATCTCTTATCTATCTATCTATCTATCTATCTATCT	2220	3241	AAAAAAATAAGAGAAAAATTTAAAAATAAATGGAACAACACTACAAGAGCTGTGTCTCTA	3300
QY	2811	TTCCCTTATGATAAGGCTCTCTCTCTCCACACACACACACACACACACACACACACACA	2870	3891	GATGAGCTACTTAGTTAGCTGATATTTTGGTATTTTAACTTTTAAAGTCAGGCTGTCTCA	3950
Db	2221	TTCCCTTATGATAAGGCTCTCTCTCTCCACACACACACACACACACACACACACACACA	2280	3301	GATGAGCTACTTAGTTAGCTGATATTTTGGTATTTTAACTTTTAAAGTCAGGCTGTCTCA	3360
QY	2871	CA	2930	3951	CCTGCACCTACATTTTAAATAATATCAATTTCTCAATGTATATCCACACAAAGCTGTGACGT	4010
Db	2281	CA	2340	3361	CCTGCACCTACATTTTAAATAATATCAATTTCTCAATGTATATCCACACAAAGCTGTGACGT	3420
QY	2931	CTCTGCCTTTCTCATCTACACAGCCAGGAGGTAAGTTATATATAAGAGGATTTATTGG	2990	4011	GAATGTTCTATAGTACCTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATC	4070
Db	2341	CTCTGCCTTTCTCATCTACACAGCCAGGAGGTAAGTTATATATAAGAGGATTTATTGG	2400	3421	GAATGTTCTATAGTACCTTTATTCACAAAACCCCAAGTAGAGACTATCCAAATATCCATC	3480
QY	2991	TAAGAGATGATGCTTAATCTCTTTAAACACTGGGCTCAAGAGAGAAATTTCTTTCTTCT	3050	4071	AACAAAGTGAACAAATAAATAAATAATCTATATCCATGCAATGGAATACCACTGCACT	4130
Db	2401	TAAGAGATGATGCTTAATCTCTTTAAACACTGGGCTCAAGAGAGAAATTTCTTTCTTCT	2460	3481	AACAAAGTGAACAAATAAATAAATAATCTATATCCATGCAATGGAATACCACTGCACT	3540
QY	3051	GTACTTTAAGCACCTATTATGTTGAGCTTATATATACAAAGGTTATATATGCTA	3110	4131	ACAAAGGAAGAGCTTACTTGGGATGAATCCCAAGTCAATGCGCTTAATGAAAGAGTCA	4190
Db	2461	GTACTTTAAGCACCTATTATGTTGAGCTTATATATACAAAGGTTATATATGCTA	2520	3541	ACAAAGGAAGAGCTTACTTGGGATGAATCCCAAGTCAATGCGCTTAATGAAAGAGTCA	3600
QY	3111	ATATAGTAATAGTAAATGTTGGTACTATGTTAAATACCATAAAATTAATATCCCTTT	3170	4191	GACATGAAGGAGGAGATTAATGTATGCCATACGAAATTTCTAGAAAATGAAAGTAACTTATA	4250
Db	2521	ATATAGTAATAGTAAATGTTGGTACTATGTTAAATACCATAAAATTAATATCCCTTT	2580	3601	GACATGAAGGAGGAGATTAATGTATGCCATACGAAATTTCTAGAAAATGAAAGTAACTTATA	3660
QY	3171	TAAATAAAGCTAATATATTTAGTATCTTTTATGATTTATTTATGTTTATGTTT	3230	4251	GTTACAGAAAACAAATCAGGCAGGCTAGAGGCTCACACCTGTATCCACGACCTTTCA	4310
Db	2581	TAAATAAAGCTAATATATTTAGTATCTTTTATGATTTATTTATGTTTATGTTT	2640	3661	GTTACAGAAAACAAATCAGGCAGGCTAGAGGCTCACACCTGTATCCACGACCTTTGA	3720
QY	3231	TTGATTTTTTAAAGACAATCTACCCCTGTTTACCCAGGCTGAGTGCAGTGCATCA	3290	4311	GAGCCAGCTGGGAGATTTGCTAGAACTCAGGAGTTCAAGACCAAGCTGGGCAACACACT	4370
Db	2641	TTGATTTTTTAAAGACAATCTACCCCTGTTTACCCAGGCTGAGTGCAGTGCATCA	2700	3721	GAGCCAGCTGGGAGATTTGCTAGAACTCAGGAGTTCAAGACCAAGCTGGGCAACACACT	3780
QY	3291	TAGCTTTCTGAGCTCTTGAATCTCTGGGCTCAAGCAATCTCTGCTTGGCTCCCAAA	3350	4371	GAACCTCAATTTCTCCACAAAATGGGAAAAAAGAAAGCAATAGTGGTGTCTGTGG	4430
Db	2701	TAGCTTTCTGAGCTCTTGAATCTCTGGGCTCAAGCAATCTCTGCTTGGCTCCCAAA	2760	3781	GAACCTCAATTTCTCCACAAAATGGGAAAAAAGAAAGCAATAGTGGTGTCTGTGG	3840
				4431	GGAGGGAGGACTGCAAGAGGGAGAAAGCTCTGTTGGGTGAGGGTGGTATTCAGCT	4490

Db 3841 GGAGGGGAAGGACTGCAAGAGAGGGAAGAGCTCTGGTGGGTGAGGGTGTGATTGAGGT 3900  
QY 4491 TCTGTATCTGACTGTGTAGCAGTTTGGGGTGTGTTTACATCCAAAAATATTCGTAGCAAT 4550  
Db 3901 TCTGTATCTGACTGTGTAGCAGTTTGGGGTGTGTTTACATCCAAAAATATTCGTAGCAAT 3960  
QY 4551 ATGCATCTTAAATGGGTGGAGTTTACTGTATGTAATTTATACCTCAATGTAAGAAAAAT 4610  
Db 3961 ATGCATCTTAAATGGGTGGAGTTTACTGTATGTAATTTATACCTCAATGTAAGAAAAAT 4020  
QY 4611 AATGTGTAGAAAAAGTTTCAATCTCTGCGCAGCAACGTTTATCAAAATTCCTGAGCCCT 4670  
Db 4021 AATGTGTAGAAAAAGTTTCAATCTCTGCGCAGCAACGTTTATCAAAATTCCTGAGCCCT 4080  
QY 4671 TTACTTTCGCAAAATCTCTGCACTTCTGCCCCCTACCATTAGTGTGACGACACTAGCTCCAC 4730  
Db 4081 TTACTTTCGCAAAATCTCTGCACTTCTGCCCCCTACCATTAGTGTGACGACACTAGCTCCAC 4140  
QY 4731 AATTTGGATPAATGCAATTTCTGGAAGAACTAGGGACAAAAATCCAGGCATCACTTTGTGCT 4790  
Db 4141 AATTTGGATPAATGCAATTTCTGGAAGAACTAGGGACAAAAATCCAGGCATCACTTTGTGCT 4200  
QY 4791 TTCAATATCAACACGCTGTAGAGCTGTGTTGCTGTCTGACGCTGCAATGGGAGACTCTTG 4850  
Db 4201 TTCAATATCAACACGCTGTAGAGCTGTGTTGCTGTCTGACGCTGCAATGGGAGACTCTTG 4260  
QY 4851 ATTTCTTTAAGAAACTTGGGTTACAGAGTATTTCCACAAATGCTATTCAAAATAGTGC 4910  
Db 4261 ATTTCTTTAAGAAACTTGGGTTACAGAGTATTTCCACAAATGCTATTCAAAATAGTGC 4320  
QY 4911 TTATGATATGCAAGACACTGTCTAGGAGCCAGAAAAACAAAGAGGAGGAGAAATCAGTCA 4970  
Db 4321 TTATGATATGCAAGACACTGTCTAGGAGCCAGAAAAACAAAGAGGAGGAGAAATCAGTCA 4380  
QY 4971 TTATGTTGAACAACATAGCAGATATTTAGATCATTTTGACTAGTTTAAAAAGCAGCAG 5030  
Db 4381 TTATGTTGAACAACATAGCAGATATTTAGATCATTTTGACTAGTTTAAAAAGCAGCAG 4440  
QY 5031 AGTACAAATACACATGCAATGCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 5090  
Db 4441 AGTACAAATACACATGCAATGCAATCAATCAATCAATCAATCAATCAATCAATCAATCAAT 4500  
QY 5091 AGACTAGAGGAATAAACACAGAACTTTAAACAGTCAATGTGCTAATAGACACTAAGTCTAAT 5150  
Db 4501 AGACTAGAGGAATAAACACAGAACTTTAAACAGTCAATGTGCTAATAGACACTAAGTCTAAT 4560  
QY 5151 TATTATATTAGACACTATGATATTGAGATTTAAATCTTTAAATCTTTAAATCTTTAAATCTTT 5210  
Db 4561 TATTATATTAGACACTATGATATTGAGATTTAAATCTTTAAATCTTTAAATCTTTAAATCTTT 4620  
QY 5211 GAGCTCTTCTATTTTCCATAGTATTCAGTTTGAATGATCAAGTATTAATCTTTCTTTCTT 5270  
Db 4621 GAGCTCTTCTATTTTCCATAGTATTCAGTTTGAATGATCAAGTATTAATCTTTCTTTCTT 4680  
QY 5271 TT 5330  
Db 4681 TT 4740  
QY 5331 GAATGGCATGAYCATGCTCACTGCAACCTCCACCTCTGGTTCACGAAAGCTGTGCG 5390  
Db 4741 GAATGGCATGAYCATGCTCACTGCAACCTCCACCTCTGGTTCACGAAAGCTGTGCG 4800  
QY 5391 CTCAGCCCTCCGGGTAGATGGGATTACAGGGGCCCCACCACACACTCGGCTCAATGTTGT 5450  
Db 4801 CTCAGCCCTCCGGGTAGATGGGATTACAGGGGCCCCACCACACACTCGGCTCAATGTTGT 4860  
QY 5451 ATTTTATAGATAGAGTGGGGTTTACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTC 5510  
Db 4861 ATTTTATAGATAGAGTGGGGTTTACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTC 4920  
QY 5511 AGAG 5514  
|||||

Db 4921 AGAG 4924

RESULT 3  
LOCUS AL160291  
DEFINITION Human DNA sequence from clone RP11-85G18 on chromosome 10, complete sequence.  
ACCESSION AL160291  
VERSION AL160291.30  
KEYWORDS GT:212111640  
SOURCE HTG.  
ORGANISM human.  
REFERENCE Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 155922)  
Dunn, M.  
Direct Submission  
Submitted (21-MAY-2002) Wellcome Trust Sanger Institute, Hinxton,  
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
humquery@sanger.ac.uk  
Clone requests: clonerequest@sanger.ac.uk  
On May 25, 2002 this sequence version replaced gi:18958078.  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30);  
an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em: EMBL; SW: SWISSPROT; Tr: TrEMBL; Wp: WORMPEP; Information on the WORMPEP  
database can be found at  
http://www.sanger.ac.uk/projects/C.elegans/wormpep  
This sequence was generated from part of bacterial clone contigs of human  
chromosome 10, constructed by the Sanger Centre Chromosome 10  
Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chr10  
RP11-85G18 is from the library RPCI-11.1 constructed by the group  
of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: pBac3.6.  
Location/Qualifiers  
1..155922  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="10"  
/clone="RP11-85G18"  
/clone\_lib="RPCI-11.1"

BASE COUNT 45636 a 32268 c 31587 g 46431 t  
ORIGIN

Query Match 4.6%; Score 251.8; DB 9; Length 155922;  
Best Local Similarity 50.3%; Pred. No. 6.4e-943; Indels 14;  
Matches 1030; Conservative 1; Mismatches 993;

QY 3539 TTAAAGTGAACAGACAGACAGCCAGGTGTGGTCTCACGCTCTAATCCAGCACTCTGGG 3598  
|||||  
Db 108116 TTTAAATTTTAAATTTGGCCAGTTGTGGTCTCATGCTCTAATCCAGCACTTTGTG 108175  
QY 3599 AGGCTGAGTGGGTGGATGGCTTTGAGCCCTCGAGTTCAAGACAGCCTGAGCAACATGCG 3658  
|||||  
Db 108176 AGGCTGAGTGGGTGGATGGCTCACCTGAGCCAGGTTCAAGGCGAGCTGGGCAACATAGC 108235  
QY 3659 AAAACCCCTGTTTCT-----ATAACAAAAATAGCCGGCATGTGGCATGTGCTGTG 3711  
|||||  
Db 108236 AAAACCCCTGTTTCTACAAATAAAAAATAGCTGGGTGTGATGGCA-CACCCCTGTA 108294



	/organism="Homo sapiens"	/note="L2 repeat: matches 1293. .1445 of consensus"
	/db_xref="taxon:9606"	13697. .13763
	/chromosome="20"	/note="L2 repeat: matches 2607. .2674 of consensus"
	/clone="RP5-1167E19"	14119. .14269
	/clone_lib="RPCI-5"	/note="L1M4 repeat: matches 4384. .4543 of consensus"
	145. .449	14432. .14650
misc_feature	/note="match: STS: Em:Z67317"	/note="L1MC/D repeat: matches 5428. .5642 of consensus"
	complement(172. .577)	14802. .15103
misc_feature	/note="match: GSS: Em:AQ729312"	/note="AlusX repeat: matches 2. .303 of consensus"
	278. .333	15307. .15725
repeat_region	/note="28 copies 2 mer ag 76% conserved"	/note="MLN1J repeat: matches 68. .507 of consensus"
	720. .756	15745. .16662
repeat_region	/note="L2 repeat: matches 2379. .2415 of consensus"	/note="L1M4 repeat: matches 3730. .4663 of consensus"
	1741. .1785	16794. .16875
repeat_region	/note="15 copies 3 mer ata 100% conserved"	/note="L1M4 repeat: matches 3460. .3538 of consensus"
	1807. .1893	16876. .17183
repeat_region	/note="MIR repeat: matches 39. .122 of consensus"	/note="Alujo repeat: matches 1. .304 of consensus"
	2601. .2693	17184. .17510
repeat_region	/note="HAL1 repeat: matches 5. .98 of consensus"	/note="L1M4 repeat: matches 3115. .3460 of consensus"
	2743. .2880	17439. .17841
repeat_region	/note="LTR45 repeat: matches 389. .525 of consensus"	/note="match: GSS: Em:B32565"
	3243. .3403	17511. .17611
repeat_region	/note="HAL1 repeat: matches 425. .559 of consensus"	/note="HAL1 repeat: matches 707. .813 of consensus"
	3404. .3703	complement(17614. .18091)
misc_feature	/note="Alusq repeat: matches 1. .300 of consensus"	/note="match: GSS: Em:AQ264589"
	complement(3553. .4081)	17689. .17892
misc_feature	/note="match: GSS: Em:B95838"	/note="L1ME3 repeat: matches 5524. .5728 of consensus"
	complement(3631. .4052)	18401. .18707
repeat_region	/note="match: GSS: Em:AQ010447"	/note="Alusg repeat: matches 1. .308 of consensus"
	3704. .3758	19010. .19193
repeat_region	/note="HAL1 repeat: matches 559. .618 of consensus"	/note="MLTII repeat: matches 3. .232 of consensus"
	3956. .4352	20676. .20803
repeat_region	/note="HAL1 repeat: matches 360. .749 of consensus"	/note="64 copies 2 mer ag 62% conserved"
	4353. .4643	21194. .21437
repeat_region	/note="AlusX repeat: matches 1. .290 of consensus"	/note="match: GSS: Em:AQ834343"
	4644. .4708	22509. .22811
repeat_region	/note="HAL1 repeat: matches 295. .360 of consensus"	/note="Alujo repeat: matches 1. .297 of consensus"
	6054. .6145	complement(23846. .24267)
repeat_region	/note="MIR repeat: matches 93. .197 of consensus"	/note="match: GSS: Em:AQ562784"
		complement(23854. .24167)



[illegible]





[illegible]



repeat_region	4565..4707	QY	3530	GGGAGACATTTAAAGTGAACACAGACAGCCAGGTGTGGTGGCTCAGCCTGTAAATCCAG	3589
repeat_region	/rpt_family="FRAM"	Db	62777	GGGGATCTTTTAAATGTGATCGCGCGGTGCGGTGCTAGACCTGTACTCCAG	62718
repeat_region	5017..5062	QY	3590	CACCTGGAGGCTGAGGTGGTGGATCGCTTGAGCCCTGAGTGTCAAGACAGCAGCTGAG	3649
repeat_region	5068..5201	Db	62717	CACCTTTGGAGGCTGAGCGGGGAGTCACTTGAAGTCAAGAGTTCAGACAGCAGCTGCG	62658
repeat_region	5245..5444	QY	3650	CAACATGGCAAAACCCCTGTTCT-----ATAACAAAAATAGCCGGCATGTTGGCATG	3703
repeat_region	5477..5767	Db	62657	CAATATGGCAAAACCCATCTCTACTAAAAACACCAAAATTAAGCTGGCGTATGGTGG	62598
repeat_region	5769..5808	QY	3704	TGCCTGTGTCCTCCAGCTA-CTAGGGGCTGAGGAGAGAAATCTTTGGAGCCAGAGGT	3762
repeat_region	5823..6485	Db	62597	TGCCTGTGTCCTCAGCTACCTGGGAAGTTGAGCCATGAGGAATGATTGAACAGGAGGT	62538
repeat_region	complement(6495..6787)	QY	3763	CAAGCTGCACCTGAGCAGTGTGGGCCACTGCACTCCAGCTCCAGCTGGGTGACAGACAGAC	3822
repeat_region	complement(6789..6914)	Db	62537	TAGGCTGCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA	62478
repeat_region	complement(7100..7282)	QY	3823	CTTGCTCAAAAAATTAAGAGAAAAATTAATAATAATAATAATAATAATAATAATAATA	3882
repeat_region	complement(7424..7587)	Db	62477	TCTGTCTCAAAAAATTAATAATAATAATAATAATAATAATAATAATAATAATAATAATA	62418
repeat_region	7646..7695	QY	3883	TTGCTCTAGATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTTAAAGTCAGG	3942
repeat_region	/rpt_family="CA)n"	Db	62417	T-----CACTCCAGCCATTCGAAATTCGCCAGGTGATGCTACATGCAGTAAAGCTGG	62362
repeat_region	8500..8803	QY	3943	GTCGTGCACCTGACATGATTAATAATAATAATAATAATAATAATAATAATAATAATA	4002
repeat_region	9439..9743	Db	62361	AAGCATGCACATGCTAAGGTTCCTGGCGCGAGCCGTCATATAATAATAATAATAATA	62302
repeat_region	complement(9831..10142)	QY	4003	TGGTACGTGAATGTTTCATAGTACCTTTTATCACAAACCCCAAACTAGAGACTATCC	4062
repeat_region	10143..10172	Db	62301	AAGTATACAGACACCCAGATCTCTTTTCAAAATCTCCTGAGTTGGTCT-----CCAGG	62247
repeat_region	/rpt_family="AT-rich"	QY	4063	TATCCATCAACAAGTGAACAAATTAACAAATGTCTATATCCATGCATGGAATACCA	4122
repeat_region	10519..10697	Db	62246	TATCTACCTTTTTCACAGGCTTTCAGTAATCTCGAGAGCAGCTGCGGTGCTGCTACT	62187
repeat_region	complement(10740..10867)	QY	4123	CCTGAGTACAAAGAAAGCAAGTCTCTGGGATGAATCCCAAGTCAAGTCAAGCCTTA	4182
repeat_region	10878..11176	Db	62186	CATCCAGCATAACATAAGGCAATGTTTGTGGCGTGGTGGTGGTGGTGGTGGTGGTGG	62127
repeat_region	/rpt_family="PAM"	QY	4183	AAGATCAGACATGAAGGAGGAGATAATGATGCCATAGCAATCTAGAAAT---GAA	4239
repeat_region	11585..11717	Db	62126	GACTGGGTGAATTTGAGGACTTCCATAGGCCCTCCTTTATCCATTTGGGTTTCACTGT	62067
repeat_region	11721..12039	QY	4240	AGTAACTTATAGTTACAGAAAGCAATCAGGCGAGGCATAGAGGCTCACACCTGTAATCC	4299
repeat_region	complement(12724..13037)	Db	62066	GGGATGTTCATGAAACTATTAGTTAGCCGGGCGATGGTGGCTCAGCCCTGTAATCC	62007
repeat_region	/rpt_family="AluSp"	QY	4300	CAGCACTTTGAGAGGCC--ACGTGGGAAGATTGCTAGAACTCAGGAGTTCAGACAGCCT	4358
repeat_region	complement(14362..14412)	Db	62006	TAGCACATTTGGAGGCCAAGCAGGTGGATCACCTCAGTCAAGTTCAGAGTTCGAGACC	61947
repeat_region	complement(14413..14703)	QY	4359	GGGCAACACAGTGAACCTCAATTCACACAAAAATGGGAAAAAAGAGCAAAATCAGTG	4418
repeat_region	complement(14704..14797)	Db	61946	TGCCAACATGGCAAAACCCCATCTCTACTAAAAATACAAAAATAGCCAGGATGGTGT	61887
repeat_region	14801..14998	QY	4419	GTGTGCTGTGGGAGGGGAA---GGACTGCAAGAGAGGAGAAAGCTCTGGTGGGTGAG	4475
repeat_region	/rpt_family="L1MC4a"	Db	61886	GGGCACCTGTAAATCCAGCTACTTGGAGGCTGAGCGGGAGAAATCACTGGAACCCGAG	61827
repeat_region	complement(15118..15351)	QY	4476	GGTGTGATTCAGGTTCTGTATCCCTGACTGTGGTGTAGCAGTTTGGGTGTTTACATCCA	4535
repeat_region	15396..15690	Db	61826	GCTGGAAGTTGCAAGTGAAGGAGATCACACCAATTTTCACTCCAGCCTTGGTGAAGAG	61767
repeat_region	/rpt_family="AluSg"	QY	4536	AAATATTCGTAGAAATTAATGCATCTTAAATGGGTGGAGTTTACTGTATGTAATAATAC	4595
repeat_region	complement(15701..15982)	Db	61766	GACCCCGCTCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAA	61707
repeat_region	16311..17324	QY	4596	AATGTAAGAAAAAATTAATGTGTAAGAAAAAGTTTCAATTTCTTTGCCAGCAACGTTAT	4655

Db	61706	GACCAAAATGATCTGGCAATTGATATTTTGGATGATTTAAATTAATCAACTCACTAG	61647
QY	4656	AAATTCCTGAGCCCTTTACTTCGCAAAATCTCTGCACTTCTGCCCCGTACCATTAGGTGA	4715
Db	61646	CAGACAGTGTTCCTGTAGACAAATCGCTTTTGTAAATTTTACGTGAGCAAGATCA	61587
QY	4716	CAGCACTAGCTCCACAAATTTGGATAAATGCATTTCTTGAAAAAGACTAGGACAAATCCA	4775
Db	61586	GAGTTTAAAGTTTACAAGTCACCGCAACAAGTAATAGGATGACAGTAGGTTCAGATCCCA	61527
QY	4776	GCATCAGCTTGCCTTCATATCAACCAAGCTGTACAGCTTGTGTTGCTGTCTGCGCTG	4835
Db	61526	-----TAAAGAGAGCGGAGGTGGAGCTTGGAACTCACCATTAGGCTTAAG	61483
QY	4836	CAATGGGGACTCTGATTTCTTTAAGGAAACTTGGGTTACCAGAGATTTTCCACAAATGC	4895
Db	61482	TATCTTGATCTCTTAACCTAGTGGGAGGCGCACAGGTACAAAATCTATGTTCAATTA	61423
QY	4896	TATCAAAATAGTCTTATGATATGCAAGACACTGTGCTAGGAGCGCAAAAACAAGAGG	4955
Db	61422	TTTCCCTTACTTATATGAATACTTAATTTCCCTTTAATCTTTTGTAGATTGCAAAAGTA	61363
QY	4956	AGGAGAAATCAGTCAATATGTTGGGAACACATAGCAAGATATTTAGATCTTTGACTAG	5015
Db	61362	ATATGAGCTCAATGTAGCTAATGAACA--ATGCCAGGATATTTTCTCTCACCGCCCC	61305
QY	5016	TTAAAAAGCAGCAGAGTACAAATCACACATGCAATCAGTATATCCAAATCATGTAA	5075
Db	61304	CTCCAGCCGCCAGCTCTCCGCCAGTGACCAATATTACAGCAATCTCATTTCAATTTTA	61245
QY	5076	TATGTGCCCTGTAGAAGACTAGAGGAATAAACA---CAAGAATCTTAACAGTCAATGTG	5131
Db	61244	GTTAGAAATGAGTTGTTCTACACAAATGAATTTGTCCAAGATCAAAACTGTTCTTTC	61185
QY	5132	ATTAGACACTAGTCTAATTAATTTATTAGACACTATGATATTTGAGATTTAAAAAATC	5191
Db	61184	ACTTATGTTTGAATACATGATCAGCAATAACTAGCAATAAATTTCTCCAGT---AGCAG	61128
QY	5192	TTTAATATTTTAAATTTAGAGCTCTCTATTTTCCATAGTATTCAGTTTGACAATCA	5251
Db	61127	TTTTCTCAGATTAATTCACATCCCGAGTTTACCCCTAATTCCTGAAGTCTGCCCCAGG	61068
QY	5252	TCAGTATPACTCTTCTT- -TTTTTTTTTTTTTTTTTTTTTTTGGATGGAGTTTGTG	5309
Db	61067	CAATGGTTGCTCACAAGTCAGGGTTATTATTATTTTTTTTTTTTGAGACGAAGTTCACT	61008
QY	5310	CTTGTGCCCAGTCTGGAGTGAATGGCATGAYCATAGTCACTGCAACCTCCACTCCT	5369
Db	61007	CTTGTGCCCAGGCTGGAGTGAATGGCGCATTTTGTCTACCAACACCTCCACTCCC	60948
QY	5370	GGGTTCAAGCAAGCTGTGCGCTCAGC--CTCCCGGTAGATGGGATTACAGCGGCCAC	5427
Db	60947	AGTTCAAGTATCTCTCGCTCAGCTCTCCCAAGTAGCTGGGATCACAGCGGTGTC	60888
QY	5428	CACCACACTCGCTAATGTTGATTTTTAGTAGAGATGGGTTTACCATGTTGGCCAG	5487
Db	60887	CACCATGCTGCTAAT- -TTTGTATTTTGTAGTAGATGGGTTTCTCCATGTTGTGTCAG	60829
QY	5488	GCTGCTCAAACTCTGACCTCAG	5512
Db	60828	GGTGCTCAAACTCTGACCTCAG	60804

RESULT 7  
AC005291/c AC005291 Homo sapiens chromosome 17, clone hRPK.401\_O\_9, complete sequence.  
DEFINITION AC005291  
ACCESSION AC005291  
VERSION AC005291.1 GI:3402737  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 198582)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 17, clone hRPK.401\_O\_9  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 198582)  
AUTHORS Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatlin,C., Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E., Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gartyna,S., Gensheimer,S., Geraghty,K., Gilmartin,T., Grant,G., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychalecky,J., Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C., Strickland,C., Subramanian,A., Tortuella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (17-JUL-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
REFERENCE 3 (bases 1 to 198582)  
AUTHORS Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,W., Collins,S., Collymore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P., Fitzhugh,W., Forrest,C., Funke,R., Gage,D., Gartyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczeky,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychalecky,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Tichovolsky,N., Tortuella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D., Ye,W.J., Zhao,J. and Zody,M.  
TITLE Direct Submission  
JOURNAL Submitted (07-AUG-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
COMMENT On Aug 7, 1998 this sequence version replaced gi:3399679.  
All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html.

FEATURES  
Source  
1..198582  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/clone="hRPK.401\_O\_9"  
/clone\_lib="RPCT-11 human BAC library"  
59..160  
/rpt\_family="LIMC5"  
162..429  
/rpt\_family="AluSg"  
complement(578..1122)  
/rpt\_family="MER41B"  
1123..1427  
/rpt\_family="AluY"  
complement(1428..1523)  
/rpt\_family="MER41B"  
2090..2402  
/rpt\_family="AluSx"  
2414..2715  
/rpt\_family="AluJo"  
complement(2843..2927)  
/rpt\_family="MER57-internal"

```
repeat_region 3021..3315
/rpt_family="AluSp"
complement(4012..4318)
/rpt_family="AluX"
repeat_region
complement(4485..4623)
/rpt_family="MER5B"
repeat_region
complement(4726..5032)
/rpt_family="AluX"
repeat_region
complement(5033..5095)
/rpt_family="MER5B"
repeat_region
complement(5678..5998)
/rpt_family="AluX"
repeat_region
complement(6440..6559)
/rpt_family="FLAM_A"
repeat_region
complement(6564..6866)
/rpt_family="AluSc"
repeat_region
complement(6872..7224)
/rpt_family="MLTf"
repeat_region
complement(7228..7386)
/rpt_family="AluSp"
repeat_region
complement(7409..7515)
/rpt_family="MLTf"
repeat_region
7543..7624
/rpt_family="MER91A"
repeat_region
complement(7677..7966)
/rpt_family="AluSc"
repeat_region
8535..8626
/rpt_family="purine-rich"
repeat_region
8627..8749
/rpt_family="GAAA)n"
repeat_region
8850..9244
/rpt_family="MSTD"
repeat_region
complement(9535..9818)
/rpt_family="AluJo"
repeat_region
9926..10212
/rpt_family="AluX"
repeat_region
10526..10823
/rpt_family="AluSp"
repeat_region
complement(11051..11102)
/rpt_family="MADE1"
repeat_region
11409..11719
/rpt_family="AluSp"
repeat_region
complement(11896..12214)
/rpt_family="MLTfD"
repeat_region
complement(12221..12247)
/rpt_family="(CA)n"
repeat_region
complement(12248..12515)
/rpt_family="AluX"
repeat_region
complement(12516..12739)
/rpt_family="MLTfD"
repeat_region
12918..13226
/rpt_family="AluSc"
repeat_region
14743..14837
/rpt_family="MIR"
repeat_region
complement(14890..15234)
/rpt_family="MLTfA1"
unsure 15298..15317
/notes="Single-stranded coverage."
repeat_region
complement(15312..15590)
/rpt_family="AluX"
repeat_region
complement(16329..16440)
/rpt_family="Charlie3"
repeat_region
complement(16446..16560)
/rpt_family="L2"
repeat_region
16565..16875
/rpt_family="AluSp"
repeat_region
complement(17280..17592)
/rpt_family="AluSg"
repeat_region
17597..17708
/rpt_family="L2"
repeat_region
17782..17906
/rpt_family="MER96"
repeat_region
complement(18625..18911)
```

---

```
repeat_region /rpt_family="AluX"
18915..18949
/rpt_family="L2"
repeat_region
18950..19106
/rpt_family="THE1B"
repeat_region
complement(19107..19377)
/rpt_family="AluY"
repeat_region
19378..19589
/rpt_family="THE1B"
repeat_region
19590..19760
/rpt_family="L2"
repeat_region
19956..20415
/rpt_family="MLTfG"
repeat_region
20450..20581
/rpt_family="FLAM_C"
repeat_region
21565..21745
/rpt_family="MIR"
repeat_region
21746..21807
/rpt_family="MIR"
repeat_region
21889..22006
/rpt_family="AluX"
repeat_region
22007..22055
/rpt_family="(CA)n"
repeat_region
22056..22230
/rpt_family="AluX"
repeat_region
complement(22333..22557)
/rpt_family="MLTfJ"
repeat_region
complement(22715..22824)
/rpt_family="MIR"
repeat_region
complement(22825..23109)
/rpt_family="AluSp"
repeat_region
complement(23110..23192)
/rpt_family="MIR"
repeat_region
24077..24285
/rpt_family="AluY"
repeat_region
complement(24558..24863)
/rpt_family="AluX"
repeat_region
complement(25044..25171)
/rpt_family="MLTfA1"
repeat_region
25172..25469
/rpt_family="AluSg"
repeat_region
complement(25470..25682)
/rpt_family="MLTfA1"
repeat_region
complement(25893..26185)
/rpt_family="MLTfE"
repeat_region
26194..26267
/rpt_family="purine-rich"

Query Match 4.4%; Score 241.8; DB 9; Length 198582;
Best Local Similarity 50.0%; Pred. No. 5.1e-37;
Matches 1003; Conservative 1; Mismatches 948; Indels 53; Gaps 14;

QY 3530 GGGGAGACATTTAAAGTGAACACAGACAGCAGGCTGGTGGCTCACGCCCTGTAATCCAG 3589
|||||
Db 5058 GGGGGATCTTTAAAAATGTGGATCGGCCGGGTGGCTGCAGACCTGTACTCCAG 4999

QY 3590 CACTCTGGAGCTGAGGTGGTGGATCGCTTGAGCCCTGGAGTTCGAAGACCACCCCTGAG 3649
|||||
Db 4998 CACTTTGGGAGCTGAGCGGGCAGATCACTTGAAGTCAGGAGTTCACGCCCTGGC 4939

QY 3650 CAACATGGCAAAACCCCTGTTTCT-----ATAACAAAAATTAGCCGGCATGGTGGCATG 3703
|||||
Db 4938 CAATATGGCAAAACCCATCTCTACTATAAACACCAAAATAGCCTGGCTGATGGTGG 4879

QY 3704 TGCCTGTGTCACGCTA-CTAGGGGCTGAGCAGGAGAAATCTTTGAGCCACAGGAGGT 3762
|||||
Db 4878 TGCCTGTGTCACGCTACCTGGGAAGTTGAGCCATGAGGAATGATTGAACACAGGAGGT 4819

QY 3763 CAAGGTGCACCTGAGCAGTCTTGGCCCACTGCACCTCCAGCTGGGTGACAGGACAGAC 3822
|||||
Db 4818 TAAGGTGCAGTGCAGTGCATGTAACCTTCAGCTGCATGTAACCTGCATGTAACCTGCATG 4759

QY 3823 CTTGCCTCAAAAAATAAGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 3882
```

[illegible][illegible]

## RESULT 8

AC104520/C

LOCUS

## DEFINITION

ACCESSION

VERSION

## KEYWORDS

SOURCE

ORGANISM

0  
0  
:  
:  
:  
:  
:  
:  
:

## REFERENCE

## AUTHORS

TITLE  
JOURNAL

JOURNAL  
PREFERENCE

## REFERENCE AUTHORS

[illegible]FILE  
JOURNAL

JOURNAL

## REFERENCE

## AUTHORS

**TITLE**

JOURNAL

COMMENT



```
(402. .371,370. .13); match 0.98-1.00; database searched:
est"
```

[illegible]

	Query Match	4.4%	Score 241.4;	DB 9;	Length 4107;
	Best Local Similarity	49.4%;	Pred. No. 7.2e-37;		
	Matches 98;	Conservative 1;	Mismatches 357;	Indels 47;	Gaps 12;
Qy	3539	TTTAAAGTGAACAGACAGCCAGGTGTGGTGTGGCTCACGGCTGTAAATCCACGACTCTGGG	3598		
	39024	TTAAAGAGAAAAAAGTAGTTCGGTCAGTAGCTCACGGCTATAATCCCAACGCTTTGGG			



Qy	4671	TTACTTCGCAAAATTCCTCTGCACTTCTGCCCCGTACCAATTAGGTGACAGCACTAGCTCCAC	4730
Db	37897	GTGTCTCTTCCAAAGCCGTG--CTCCCGAGACAGTGGGGTCTCCAAATGTTTTGTATGATC	37840
Qy	4731	AAATTGGATAAATGCATTTCTCTGAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCT	4790
Db	37839	ACAGGTCCTAAGAGGCTTTGATGTGTTCACTGAATGTGAGGAGGAATGAGCAAA	37780
Qy	4791	TTCCATATCAACCACGCTGTACAGCTTGTGTGCTGTCTGCAGCTGCAATGGGGACTCTTG	4850
Db	37779	AGAAGAGGCACCAATTCAGGCAAAACAAACATCAGCATCTTGA AAAACCCAGCAATTGA	37720
Qy	4851	ATTTCITTAAGGAAACTTGGGTACCAGAGTATTTCCACAATGCTATTCAATTAAGTGC	4910
Db	37719	ACTGATTAAAGTAAATCCCTGATTGCTTAAGCACAGTTCAAAAGGAGTAAAGCACAGCTTGC	37660
Qy	4911	TTATGAT--ATGCAAGACACTGTGCTAGGAGCCAGAAAACAAGAGGAGGAGAAATCAGT	4968
Db	37659	TTCAATGATATTGGGGATCTGAAGTCGATTTCTGAGAAGTTAAGATGCCATTAGACCGA	37600
Qy	4969	CATTATGTGGGAACAACATAGCAAGATATTTAGATCATTTTGACTAGTTTAAAA--AAGCAG	5027
Db	37599	CAGCAACTACTAAGAAAATACTTCCAAAAATATAGTGTAAACAATCATGAAAAGAATTAA	37540
Qy	5028	CAGAGTCAAAAATCACACTCAATCAGTATATCCAAATCATGTAAATATGTGCGCTGTA	5087
Db	37539	AAATTACACTAGAAAATATTCACTTAATGCAAAAGAAACATCAAGAAGGAATAAAGG	37480
Qy	5088	GAAGACTAGAGGAATAAACACAAGAATCTTAACAGTCATTGTCTATTAGACACTAAGTCT	5147
Db	37479	AACAAAATGAAAAACAAAAGGGAATGGCAGGTACAAATCCAACATATATCAATAACGGC	37420
Qy	5148	AAATTATTATTAGACACTATGATATTGAGATTTAAAAATCTTT--TAATATTTTAAA	5205
Db	37419	ATTAAATATGATGGATTAAACATCCAATCAAAAGACAGAAATCTCAGACTGGGATTTAA	37360
Qy	5206	ATTTAGAGCTCTTCTATTTTCCATAGTATTCAAGTTTGACAATCATCAAGTATTACHTCT	5265
Db	37359	AAAAACAAGATCCAACATATATACTATATACTGTCTCAGGGGACA-----C	37310
Qy	5266	TTCCTTTTTTTTTTTTTTTTTTTTTTTTGTGAGATGGAGTTTGGTCTTGTCGCCATGCTG	5325
Db	37309	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGAGACAGATTTCACTCTTGTCGCCAGGCTG	37250
Qy	5326	GAGTGGAAATGGCATGAYCATAGCTCACTGCAACCTCCACCTCTCGGGTTCAAGCAAAAGCT	5385
Db	37249	GAGTGGCAATGGCATGATCTTGCTCACTGCAATCTCCACCTCCCGGGTTCAAGCGATCTCT	37190
Qy	5386	GTCCGCTCAGCTCCCGGTAGATGGGATTAACGGCGCCCAACCCACACTCGGCTAATG	5445
Db	37189	CTGCCTCAGCTCTCTGAGTAGCTGGGACTACAGGCATCGCGCCACCGC-----A	37139
Qy	5446	TTTGTATTTTATAGTAGAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTG	5505
Db	37138	TGTGTATTTTATAGTAGACAAAGGTTTCTCCATGTTGTCAGCTGGTCTCAAACTCCCA	37079
Qy	5506	ACCTCAG	5512
Db	37078	ACCTCAG	37072

RESULT 10	AL590609	171073 bp	DNA	linear	HTG 26-SEP-2001
LOCUS	AL590609/c	171073 bp	DNA	linear	HTG 26-SEP-2001
DEFINITION	Homo sapiens chromosome 1 clone RP11-4M23, *** SEQUENCING IN PROGRESS ***, in ordered pieces.				
ACCESSION	AL590609	171073 bp	DNA	linear	HTG 26-SEP-2001
VERSION	AL590609.14	GI:15795484			
KEYWORDS	HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
1 (bases 1 to 171073)
Wallis, J.
Direct Submission
Submitted (25-Sep-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
On Sep 27, 2001 this sequence version replaced gi:15787779.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA4M23
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: M13; M77815; 0% of reads
Chemistry: Dye-terminator Big Dye; 108752; 99% of reads
Consensus quality: 170938 bases at least Q40
Consensus quality: 171073 bases at least Q30
Consensus quality: 171073 bases at least Q20
Insert size: 171073; sum-of-contigs
Insert size: 180783; 1.3% error; agarose-fp
Quality coverage: 13.10x in Q20 bases; sum-of-contigs Quality
coverage: 12.39x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
location/Qualifiers
1. 171073
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-4M23"
/clone_lib="RPC1-11.1"
1. 171073
/note="assembly_fragment:05881
clone_end:SP6
vector_side:right
clone_end:T7
vector_side:left"
misc_feature 45416 a 38794 c 38902 g 47961 t
BASE COUNT
ORIGIN

```

Query Match	4.4%	Score 241.2	DB 2	Length 171073
Best Local Similarity	63.8%	Pred. No. 6.8e-37		
Matches 453	Conservative 0	Mismatches 238	Indels 19	Gaps 5
QY	3175	ATAAGCTAATTATTATGGATCTTTTTAGTATTCAATTTATGCTTTTTATGTTTTTGA	3234	
Db	150354	AGCAATTCCTGGTACTTTAAATATTATTGATATTATCTATTATTATTATTATTATT	150295	
QY	3235	TTTTTTAAAAGACAATCTCACCTGTTACCCAGGCTGGAGTGCAGTGGTGCAAATCATAGC	3294	
Db	150294	TTTCTTGAGATGGAGTCTCGCTCTGTGCGCCAGGCTGGAGTGCAAATGGCAGGATCTGGC	150235	
QY	3295	TTTTCTGCAGCTTTTGAACCTCTGGGCTCAAGCAATCTCTCGCTTGGCTTCCCAAGTGT	3354	
Db	150234	TCATCGACCTCCACCTCCAAGGTTCAAGCGATTCTCTGCTTCAGCTTCCCAAGTAGC	150175	
QY	3355	TGGGA-TACAGTCTATGAGCCACCTGCATCTGGCCTAGGATGCATTTAGATTAAAAATATGCA	3413	
Db	150174	TAGGATTACAGGCATGTGCCACCATGCCAGCTAAATTTGTATTTTAGTAGAGGTGAG	150115	
QY	3414	TTTTAAATTTTAAATAATATGCTAATTTTTACCTTATATGCTATGCTACTGCTAATA	3473	
Db	150114	TTTCACCATGT-----TGGCCAGGCTGATTTGAACCTCTCAGCCTCAG-----ATGATCC	150065	
QY	3474	AATCTAGTTGCTGGCCTAAAGTTTAAAGTGCTTTTCCAATAAAGCTTCATGTACGTGAG--	3531	









```
QY 4992 AGATATTAGATCATTTTTCACCTAGTTTAAAAAGCAGCAGAGTACAAAATC-----ACAC 5045
Db 27606 CTTGGTACCTTCCATGACCTCAGATTAATATCTTTGTATCATATATTCTTCAGAGTTC 27665
QY 5046 ATGCAATAGTATAATCAATCATGTAATATGTCCTGTAGAAAGACTAGAGGAATAA 5105
Db 27666 ATTGACAACAATTTCTTCAATATTTTAAATAGAAAGACAAGTTAGTACATCTACAA 27725
QY 5106 ACACGAATCTTAAACAGTCATTTGTCATTAGACACTAAGTCTAAATATTATTATTAGACA 5165
Db 27726 CCACCTAGACTCTAAGACACAGTAGGTAACAGAACTCTTCACAATGTTTGGTAATTC 27785
QY 5166 CTATGATATTGAGATTTAAAAATCTTAAATATTTTAAATTTAGAGCTCTCTATTTT 5225
Db 27786 TTAACCAACTAGACTGTTTTCCTCCCTCTATCATTTTACATATCACTTGCCTCTATTATTA 27845
QY 5226 TCCATAGTATTCAAGTTTGACAATGATCAAGTATTACTCTTCTCT--TTTTTTTTTTTTT 5282
Db 27846 AAGTATTGATACAGCATTCATCATGTTTATCTCTTCTCTCTCTCTCTCTCTCTCTCT 27905
QY 5283 TTTTTCCTTTTGGAGTGGAGTTTGGTCTTGTGTCCTGTCCTGAGTGGAGTGGAGTGGAGT 5342
Db 27906 TTTTTCCTTTTGGAGACAGAGTCTTGCTC-TGTTGCCAGGCTGGAGTGGAGTGGAGT 27964
QY 5343 CATAGTCACTGCAACCTCCACCTCTGCTGGTTCAGCAAGAGCTGTCGCCCTCAGCCTCCG 5402
Db 27965 TTTGCTCTCACTGCAACCTCCGCTCTCTGCTGGTTCAGCGATTCATGCTGCCCTCTTG 28024
QY 5403 GGTAGTGGGATTACAGCGCCGCCACACACACTCGGCTAATGTTGTATTTTATTAGTACA 5462
Db 28025 AGTACTGGGACTACAGAGTGTACACACACTGGCTGGTATTTTATATTTTATTAGTACA 28084
QY 5463 GATGGGGTTTACCATTGTTGCGGAGGCTGGTCTCAAACTCCCGACCTCAG 5512
Db 28085 GATGGGGTTTACCATTGTTGTTCAAACTGGTCTCAAACTCCCTGACCTCAG 28134

RESULT 13
AC023132
LOCUS AC023132 186107 bp DNA linear HTG 17-AUG-2000
DEFINITION Homo sapiens chromosome 8 clone RP11-421C14, WORKING DRAFT
SEQUENCE, 31 unordered pieces.
ACCESSION AC023132 GI:9838192
VERSION AC023132.4 HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 186107)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 186107)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (08-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Aug 17, 2000 this sequence version replaced gi:8439998.
```

```
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0421C14
----- Summary Statistics -----
Sequencing vector: M13; 100%
Sequencing method: plasmid; 0%
Chemistry: Dye-terminator ET; 100% of reads
Assembly program: Phrap; version 0.990319
```

```
Consensus quality: 177203 bases at least Q40
Consensus quality: 179056 bases at least Q30
Consensus quality: 180076 bases at least Q20
Insert size: 196000; agarose-fp
Insert size: 184314; sum-of-contigs
Quality coverage: 4.17 in Q20 bases; agarose-fp
Quality coverage: 4.48 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 31 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1251: contig of 1251 bp in length
* 1252: gap of unknown length
* 1352: contig of 1531 bp in length
* 2882: gap of unknown length
* 2983: contig of 1561 bp in length
* 4544: gap of unknown length
* 4644: contig of 1314 bp in length
* 5957: gap of unknown length
* 6057: gap of unknown length
* 7885: contig of 1628 bp in length
* 7886: gap of unknown length
* 9274: contig of 1489 bp in length
* 9374: gap of unknown length
* 11213: contig of 1839 bp in length
* 11313: gap of unknown length
* 12577: contig of 1264 bp in length
* 12578: gap of unknown length
* 13678: contig of 1978 bp in length
* 14655: gap of unknown length
* 14755: gap of unknown length
* 17774: contig of 3019 bp in length
* 17775: gap of unknown length
* 17875: contig of 2726 bp in length
* 20601: gap of unknown length
* 20701: contig of 4094 bp in length
* 24795: gap of unknown length
* 24895: contig of 3721 bp in length
* 28616: gap of unknown length
* 28716: contig of 4449 bp in length
* 33165: gap of unknown length
* 33265: contig of 4563 bp in length
* 37827: gap of unknown length
* 37927: contig of 5702 bp in length
* 43629: gap of unknown length
* 43730: contig of 7082 bp in length
* 50811: gap of unknown length
* 50912: contig of 4987 bp in length
* 55998: gap of unknown length
* 58369: contig of 2371 bp in length
* 58370: gap of unknown length
* 64837: contig of 6368 bp in length
* 64937: gap of unknown length
* 71657: contig of 6720 bp in length
* 71757: gap of unknown length
* 79804: contig of 8047 bp in length
* 79805: gap of unknown length
* 87292: contig of 7388 bp in length
* 87392: gap of unknown length
* 93566: contig of 6174 bp in length
* 93666: gap of unknown length
* 101493: contig of 7827 bp in length
* 101593: gap of unknown length
* 110504: contig of 8911 bp in length
* 110505: gap of unknown length
* 123442: contig of 12838 bp in length
* 123542: gap of unknown length
* 137014: contig of 13472 bp in length
* 137015: gap of unknown length
* 152690: contig of 15576 bp in length
```

\* 152691 152790: gap of unknown length  
\* 152791 169160: contig of 16370 bp in length  
\* 169161 169260: gap of unknown length  
\* 169261 186107: contig of 16847 bp in length.

FEATURES

Location/Qualifiers  
1..186107  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/clone="RP11-421C14"  
1..1251  
/note="assembly\_name:Contig9"  
1352..2882  
/note="assembly\_name:Contig15"  
2983..4543  
/note="assembly\_name:Contig16"  
4644..5957  
/note="assembly\_name:Contig17"  
6058..7685  
/note="assembly\_name:Contig18"  
7786..9274  
/note="assembly\_name:Contig20"  
9375..11213  
/note="assembly\_name:Contig21"  
11314..12577  
/note="assembly\_name:Contig22"  
12678..14655  
/note="assembly\_name:Contig23"  
14756..17774  
/note="assembly\_name:Contig24"  
17875..20600  
/note="assembly\_name:Contig25"  
20701..24794  
/note="assembly\_name:Contig26"  
24895..28615  
/note="assembly\_name:Contig27"  
28716..33164  
/note="assembly\_name:Contig28"  
33265..37827  
/note="assembly\_name:Contig29"  
37928..43629  
/note="assembly\_name:Contig30  
clone\_end:T7  
vector\_side:left"  
43730..50811  
/note="assembly\_name:Contig31"  
50912..55898  
/note="assembly\_name:Contig32"  
55999..58369  
/note="assembly\_name:Contig33"  
58470..64837  
/note="assembly\_name:Contig34"  
64938..71657  
/note="assembly\_name:Contig35"  
71758..79804  
/note="assembly\_name:Contig36"  
79905..87292  
/note="assembly\_name:Contig37"  
87393..93566  
/note="assembly\_name:Contig38"  
93667..101493  
/note="assembly\_name:Contig39"  
101594..110504  
/note="assembly\_name:Contig40"  
110605..123442  
/note="assembly\_name:Contig41"  
123543..137014  
/note="assembly\_name:Contig42"  
137115..152690  
/note="assembly\_name:Contig43"  
152791..169160  
/note="assembly\_name:Contig44"  
169261..186107

BASE COUNT 54061 a 37551 c 36981 g 54477 t 3037 others  
ORIGIN  
Query Match 4.4%; Score 240.4; DB 2; Length 186107;  
Best Local Similarity 66.2%; Pred. No. 9.6e-37;  
Matches 441; Conservative 0; Mismatches 211; Indels 14; Gaps 6;  
QY 3226 TGTGTTTATGTTTTTAAAGACAAATCTCACCTGTGTACCCAGGCTGGAGTGGAGTGGTGC 3285  
Db 98863 TGATTATGATTGACTGAGACAGTCACTCTCTGCGCCACACTGAGTGGAGTGGAGTGGTGT 98922  
QY 3286 AATCATAGCTTTCTGCAGTCTTGAACCTCTGGGCTCAAGCAATCTCTGCTTCCCTTGGCCCTC 3345  
Db 98923 GATAGTAGCTCACTGCAGTCTCCAACCTCTCTGGGCTCAAGTGTCTCTCTACCTTGGCTTC 98982  
QY 3346 CCAAGTGTGGGA-TACAGTCAATGAGCCACTGCATCTGGCTTGGCTTAGGATC---CATTTAGA 3401  
Db 98983 CCAAGTGTGGGACACAGGTATGAACACTGTGCTTGGCTTGGCTTGGCTTGGCTTGGCTT 99042  
QY 3402 TTAAAAATATGCACTTTTAAATTTTAAATTAATATGCTAATTTTACCTTTATGTATGTGT 3461  
Db 99043 AAACCTTTATGCTGTGTAATTTTATCTCACTGCTGTTTAAAGGTAAGAGTAAAGGAGT 99102  
QY 3462 ATACTGGTAATAAATCTAGTTTGC-TGCCCTAAAGTTTAAAGTCTTTTCCAATAAGCTTCA 3520  
Db 99103 GTAGAGTTACCAGTTTCAATGTACATTTAACAGGAGTTTAAATATAAGGAAATGGAGATA 99162  
QY 3521 TGTACGTGAGGGGAGACATTTAAAGTGAACAGACAGCAGGCTGGTGGCTCACCCCTG 3580  
Db 99163 ATGATGGGAAAAATAACAGTTAAAAAGAAAGTAGGAGCGCGGCATGGTGGCTCATGCCCTA 99222  
QY 3581 TAATCCAGCACTCTGGGAGGCTGAGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3640  
Db 99223 TAATCCCTGCACTTTGGGAGGCTGAGGAGGCGGCTGAGGAGGCTGAGGAGGCTGAGGAGG 99282  
QY 3641 CAGCCTGAGCAACATGGCAAAACCTGTTTCT-ATACAAAAAATAGCCCGGCA 3693  
Db 99283 CAGCCTGGCCACATGGTGAACCCCTGCTCTACTGAAATACAAAAATAGCCAGGTG 99342  
QY 3694 TGGTGGCATGTCCTGTGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3752  
Db 99343 TGGTGGCATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 99402  
QY 3753 CCCAGGAGTCAAGGCTGCACTGAGCAGTCTTGGCCACTGCCTCCAGCTCCAGCTCCAGCTCC 3812  
Db 99403 CCTGGGAGGAGAGCTGTCATGAGTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 99462  
QY 3813 AGGA-CCAGACCTTGCCTCAAAAAAATAAGAGAAAAATTAATAATGAAACAACT 3871  
Db 99463 AAGAGTGAATTCATCTCAAAAAAACAACAAACAAACAAACAAACAAACAAACAAACAAAC 99522  
QY 3872 ACAAAG 3877  
Db 99523 AGAAAG 99528  
RESULT 14  
AC113194  
LOCUS  
DEFINITION  
Homo sapiens chromosome 8 clone RP11-1070A24 map 8, \*\*\* SEQUENCING  
IN PROGRESS \*\*\*, 2 ordered pieces.  
ACCESSION  
AC113194  
VERSION  
AC113194.7 GI:22417385  
KEYWORDS  
HTG; HTGS\_PHASE2; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
1 (bases 1 to 208164)  
AUTHORS  
Birren,B., Nusbaum,C. and Lander,E.  
TITLE  
Homo sapiens chromosome 8, clone RP11-1070A24  
JOURNAL  
Unpublished

REFERENCE  
AUTHORS

2 (bases 1 to 208164)  
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,  
Anderson,S., Barna,N., Bastien,V., Boguslavskiy,L., Boukhgalter,B.,  
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,  
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cook,A.,  
Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,  
Ferrelira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,  
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-pierre,N.,  
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,  
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,  
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,  
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,  
McEwan,P., McKernan,K., Meldrim,J., Meneus,L., Mihova,T.,  
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,  
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,  
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,  
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,  
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,  
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,  
Strauss,N., Subramanian,A., Talamas,J., Testfaye,S., Theodore,J.,  
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,  
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Submitted (26-FEB-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 208164)

Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhgalter,B.,  
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collamore,A.,  
Cook,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Faro,S., Ferrelira,P., FitzGerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,  
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Testfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
COMMENT

Submitted (22-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 22, 2002 this sequence version replaced gi:22123385.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L23159

Center clone name: 1070\_A\_24

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
\* 1 41535: contig of 41535 bp in length  
\* 41536 41635: gap of 100 bp  
\* 41636 208164: contig of 166529 bp in length.

FEATURES  
source

Location/Qualifiers  
1..208164  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"

BASE COUNT 58228 a 41625 c 43295 g 64916 t 100 others  
ORIGIN

Query Match 4.4%; Score 240.4; DB 2; Length 208164;  
Best Local Similarity 66.2%; Pred. No. 9.5e-37;  
Matches 441; Conservative 0; Mismatches 211; Indels 14; Gaps 6;  
QY 3226 TCTTTTGTGATTTTAAAGACAACTCTCACCTGTGTACCAGCTGAGTGGAGTGGTGC 3285  
DB 79098 TGATTATGATGACTGAGACAGAGTCAGTCTCTGTGCCAGACTGGAGTGCAGTGGTGT 79157  
QY 3286 AATCATAGCTTTCTGCAGTCTTTGAACCTCTGGGCTCAAGCAATCCCTCTGCCCTTGGCCCTC 3345  
DB 79158 GATAGTAGCTCACTGCAGTCTCCAACCTCTGGGCTCAAGTGATCTCTTACCTTGGCTTC 79217  
QY 3346 CCAAGTGTGGGA-TACAGTCATGAGCCATGTCATCTGCCCTAGGATC---CATTTAGA 3401  
DB 79218 CCAAGTGTGGGACCAACAGGTATGAACACACTGTGCCTGGCCTGTCTCTTTAAATGAGT 79277  
QY 3402 TTAAATATGTCATTTTAAATTTTAAATTAATATGTCCTAATTTTACCTTTATGTAATGTGT 3461  
DB 79278 AAACCTTATGGTGTGTAATTTTATCTCAATCTCTGTTAAAGTAAAGAGTAAGGAGT 79337  
QY 3462 ATACTGGTAAATAATCTAGTCTTTC-TGCCCTAAAGTTTAAAGTGTCTTCCAAATAAGCTTCA 3520  
DB 79338 GTAGAGTTTACCAGTTTCAATGTACATTTAACAGGAGCTTTTAATAATAAGGAAATGGAGATA 79397  
QY 3521 TGTACCTGAGGGAGACATTTAAAGTGAACACAGACAGCCAGGTGTGGTGGCTCACGCCCTG 3580  
DB 79398 ATGATGGGAAAAATAACAGTTTAAAGAGAAAGTAGGAGGCCGGGCATGTGGCTCATGCCCTA 79457  
QY 3581 TAATCCAGCACTCTGGGAGGCTGAGGTGGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 3640  
DB 79458 TAATCCCTGCACCTTGGGAGGCTGAGGAGGCGGATCACCTGAGGTGACAGGAGTTCAGAC 79517  
QY 3641 CAGCCTGAGCACATCGGCAAAACCCCTGTTTCT-----ATAACAAAAATTAGCCGGGCA 3693  
DB 79518 CAGCCTGGCAACATGCTGAAACCCCTGCTCTACTGAAAAATACAAAAAATTAGCCAGGTG 79577  
QY 3694 TGGTGGCATGTGCTGTGCTGCCAGTACT-AGGGGGCTGAGGAGGAGAGTCTTTGGAG 3752  
DB 79578 TGGTGGCATGTGCTGTGTAATCCAGCTACTCAGGAGGCTGAGGAGGAGAGTCACTTTGAA 79637  
QY 3753 CCAGGAGGTCAAGGCTGCAGTGCAGTGTGCTTGGCCCTGACCTCCAGCTCCAGCTGGGTGAC 3812  
DB 79638 CCTGGGAGGACAGCTTGCATGAGCTGAGATGCGCACCTGCACCTCCAGCTGGGCAAC 79697  
QY 3813 AGGA-CCAGACCTTGCTCTCAAAAAATAACAGAAAAATTTAAAAATAAATGGAACAACCT 3871  
DB 79698 AAGAGTGAATTCATCTCAAAAAACAAAAACAAAAACAAAAACAAAAACAAAAACAAAAA 79757  
QY 3872 ACAAAG 3877  
DB 79758 AGAAAG 79763

## RESULT 15

AC069335  
LOCUS AC069335  
DEFINITION Homo sapiens BAC clone RP11-723C11 from 7, complete sequence.  
ACCESSION AC069335  
VERSION AC069335.17 GI:15290496  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 155521)  
Sulston, J.E. and Waterston, R.  
Toward a complete human genome sequence  
Genome Res. 8 (11), 1097-1108 (1998)  
99063792  
9847074

2 (bases 1 to 155521)  
Ozersky, P., Du, H. and Maupin, R.  
The sequence of Homo sapiens BAC clone RP11-723C11  
Unpublished (2001)  
3 (bases 1 to 155521)  
Waterston, R.H.  
Direct Submission  
Submitted (27-MAY-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
4 (bases 1 to 155521)  
Waterston, R.H.  
Direct Submission  
Submitted (25-AUG-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
5 (bases 1 to 155521)  
Waterston, R.H.  
Direct Submission  
Submitted (26-AUG-2001) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
6 (bases 1 to 155521)  
Waterston, R.  
Direct Submission  
Submitted (09-JAN-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Aug 25, 2001 this sequence version replaced gi:13431253.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@watson.wustl.edu](mailto:sapiens@watson.wustl.edu)  
----- Summary Statistics  
----- Center project name: H\_NH0723C11  
-----

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send [mailto:egreen@nhgri.nih.gov](mailto:mailto:egreen@nhgri.nih.gov), or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatenio, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome

libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)  
VECTOR: pBACe3.6  
NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-785H2; the clone sequenced to the right is RP4-592P3, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-723C11; actual end is at base position 46589 of RP4-592P3.

Polymorphisms exist between AC069335 and AC006452.  
Location/Qualifiers  
1. .155521  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7"  
/clone="RP11-723C11"  
/clone\_lib="RPCI-11"  
18. .264  
/note="match to EST AW814659 (NID:g7907653)"  
misc\_feature  
repeat\_region  
285. .286  
/rpt\_family="(TTTGTG)n"  
572. .854  
/rpt\_family="Alu"  
805. .810  
/note="match to EST BG721725 (NID:g14000912)"  
repeat\_region  
878. .1170  
/rpt\_family="Alu"  
1210. .1505  
/rpt\_family="Alu"  
1508. .1818  
/rpt\_family="Alu"  
1789. .1837  
/rpt\_family="A-rich"  
1838. .1924  
/rpt\_family="L1"  
1925. .2100  
/rpt\_family="Alu"  
2093. .2130  
/rpt\_family="(T)n"  
2101. .2417  
/rpt\_family="Alu"  
2418. .2549  
/rpt\_family="Alu"  
2550. .2609  
/rpt\_family="L1"  
2599. .2622  
/rpt\_family="AT-rich"  
2616. .2709  
/rpt\_family="L1"  
2911. .3200  
/rpt\_family="Alu"  
3237. .3454  
/rpt\_family="MER1\_type"  
3455. .3606  
/rpt\_family="MER1\_type"  
3662. .3850  
/note="match to EST BF327818 (NID:g11298566)"  
misc\_feature  
3665. .3850  
/note="similar to Bos taurus EST AW660583 (NID:g7426410)"  
3685. .3850  
/note="similar to Mus musculus EST AA667697 (NID:g2626398)"  
vv30g10.r1  
3963. .4256  
/rpt\_family="Alu"  
4303. .4459  
/rpt\_family="Alu"  
4317. .4343  
/note="similar to Homo sapiens EST BF978105 (NID:g12345320)"  
4320. .4341  
repeat\_region







```

Db      15259  TCCATGCACTATGGAATACACTCAGCAATAAAACTACTAGTGGTATACAGAAAAACA 15318
QY      4950  AAGAGGAGGAGAAATCAGTCATATATGTGGAAACAACATAGCAAGATATTTAGATCATTTT 5009
Db      15319  AACTACTGATTATACACAACACTTGAATGGATCTCAGAGGATTTATGCTGAGTGAAAAA 15378
QY      5010  GACTAGTTAAAAAGCAGAGAGTACAAAATCACACATGCAATCAGTATAATCCAAATCA 5069
Db      15379  GACAATCTCAAAAGGTTACATATATGAATCCATTTATATAATTTCTGGAAGTGAATAG 15438
QY      5070  TGTAATATGTGCCCTGTAGAAGACTAGAGGAATAACACACAGAATCTTACAGTCATTG 5129
Db      15439  ATGTGTGTAGCAGGGGGAGGACCGGGGGTGTGAGAGTTATAAGGAATAGCATGA 15498
QY      5130  TCATTAGACACTAAGTCTAAATATTATTATTAGACACTATGATATTTGAGATTTAAAAAA 5189
Db      15499  AAGACAGCTTTCTGAGAAATAGAACTGTTTCAGTGCCTTGACTGTGGTGGTATGAAGGA 15558
QY      5190  TCT-TTAATATTTTAAAAATTTAGAGCTCTCTCTATTTTCCATAGTATTCAAGTTTGACA 5248
Db      15559  TCTACGCATGTGATCAAAATTTGTATATAATTACAAAACATACACAAAATAAGTGTGGGA 15618
QY      5249  TGATCAAGTATTACTCTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 5308
Db      15619  CATTTATTCAATTCATGACAAAGAACTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTT 15678
QY      5309  TCTTGTGCCCATGTGGAGTGGAAATGGCATGAYCATAGCTCACCTGCAACCTCCACCTCC 5368
Db      15679  TCTTGTACCAGGCTGGAGTGCAGTGGCAACATCTTAGTTCACCTGCAACCTCCGCCCTCC 15738
QY      5369  TGGGTTCAAGCAAAAGCTGTCCGCTCAGCCTCCCGGTAGATGGGATTACAGCGCCCAACC 5428
Db      15739  CGGGTTCAAGTGATTCCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGCGCGTGCC 15798
QY      5429  ACCACACTCGGCTAA-TGTTTGTATTTTAGTAGAGATGGGGTTTTCACCATGTTGGCCAG 5487
Db      15799  ACCATACCCAGCTAAATTTTGTGTGTTTGTAGTAGAGCGGGGTTTTCACCATGTTGGCTGG 15858
QY      5488  GCTGGTCTCAAACTCCTGACCTCAG 5512
Db      15859  GATGCTCTCAAAATTCCTGACCTCAG 15883

```

Search completed: June 19, 2003, 07:46:51  
 Job time : 14099.9 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 20:16:19 ; Search time 1128.86 seconds  
(without alignments)  
11000.051 Million cell updates/sec

Title: US-09-966-880A-9  
Perfect score: 5514  
Sequence: 1 acagacgaatacatagtgcca.....tcaaactcctgacctcagag 5514

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues 4370478  
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*  
1: /SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.\*  
23: /SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.\*  
24: /SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5512.8	100.0	5514	21 AAC55313	Human activation-i
2	4922.8	89.3	11204	21 AAC55339	Human activation-i
3	231.6	4.2	112460	24 ABK83567	Human CDNA differe
C 4	230.2	4.2	28818	22 AAL35901	Human musculoskele
C 5	227.8	4.1	112460	24 ABK83567	Human CDNA differe
C 6	227.6	4.1	4126	22 AAK80162	Human immune/haema
C 7	227.6	4.1	4126	22 AAK89461	Human digestive sy
C 8	226.8	4.1	32204	22 AAS39620	Genomic sequence #
C 9	226.8	4.1	32204	22 AAK89019	Human digestive sy

10	226.8	4.1	32204	22 AAK91533	Human digestive sy
11	226.8	4.1	32204	22 AAI57790	Human colorectal c
12	225.4	4.1	28588	22 AAK89418	Human digestive sy
13	225.4	4.1	28588	22 AAS31923	Human liver associ
14	225.4	4.1	28588	24 ABN90278	Human liver antige
15	221.6	4.0	10855	22 AAL05277	Human reproductive
16	221.6	4.0	10855	23 ABL98161	Human testicular a
C 17	221.4	4.0	10736	22 AAK80161	Human immune/haema
C 18	221.4	4.0	10736	22 AAK89460	Human digestive sy
C 19	221.2	4.0	32152	22 AAS39621	Genomic sequence #
C 20	221.2	4.0	32152	22 AAK89020	Human digestive sy
21	221.2	4.0	32152	22 AAK91534	Human digestive sy
22	221.2	4.0	32152	22 AAI57791	Human colorectal c
C 23	221.2	4.0	36933	22 AAK66362	Human immune/haema
C 24	218.8	4.0	32134	22 ABA15354	Human nervous syst
C 25	218.8	4.0	32134	22 ABA15813	Human nervous syst
C 26	218.8	4.0	32134	22 ABA15813	Human reproductive
C 27	217.8	3.9	23241	22 AAF97870	Human neuroblastom
C 28	217.8	3.9	23241	22 AAF97871	Human neuroblastom
29	216	3.9	121162	21 AAF66548	Human kinesin-like
30	215.8	3.9	154465	24 AAD28763	Human AKAP allelic
31	215.8	3.9	158245	24 AAD28762	Human AKAP allelic
32	215.8	3.9	161425	22 AAH02340	Human AKAP10 gene
33	215.8	3.9	162025	22 AAH02339	Human AKAP10 gene
34	215.8	3.9	162025	24 AAD28758	Human AKAP allelic
35	215.8	3.9	162025	24 AAD28759	Human AKAP allelic
C 36	215	3.9	19472	22 AAS26724	Human genomic DNA
C 37	214.4	3.9	17239	22 AAS38826	Human cardiovascular
C 38	213.8	3.9	14049	22 ABA15812	Human nervous syst
39	213.6	3.9	32195	22 AAS36105	Human cardiovascular
40	213.6	3.9	32195	22 AAS31532	Human DNA for a no
41	213.6	3.9	32195	22 AAS31538	Human DNA for a no
42	213.6	3.9	32195	23 ABK44045	Genomic DNA encodi
43	213.6	3.9	32195	24 ABQ66856	Human polynucleoti
44	213.6	3.9	32195	24 ABQ66862	Human polynucleoti
C 45	212.8	3.9	33147	22 AAK67282	Human immune/haema

ALIGNMENTS

RESULT 1

AAC55313

ID AAC55313 standard; DNA; 5514 BP.

XX AAC55313;

AC AAC55313;

DT 05-FEB-2001 (first entry)

XX Human activation-induced cytidine deaminase genomic DNA SEQ ID NO:9.

XX Activation-induced cytidine deaminase; AID; cytidine deaminase;  
KW immune related disease; allergy; allergic disease; anti-allergic;  
KW antianemic; antisthmatic; ophthalmological; anti-HIV; dermatologic;  
KW gene therapy; B cell associated immune system disorder; food allergy;  
KW immunodeficiency disease; immunoglobulin A deficiency disease; asthma;  
KW IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;  
KW drug allergy; allergic rhinitis; Rosen disease; Digeorge disease; AIDS;  
KW ataxia telangiectasia; common variable immunodeficiency disorder;  
KW major histocompatibility class II deficiency disease;  
KW auto immunodeficiency syndrome; IgG subclass selection disorder; ds.

XX Homo sapiens.

OS Homo sapiens.

XX WO200058480-A1.

XX 05-OCT-2000.

XX 28-MAR-2000; 2000WO-JP01918.

XX 29-MAR-1999; 99JP-0087192.

XX 24-JUN-1999; 99JP-0178999.

XX 27-DEC-1999; 99JP-0371382.



Db	2701	GAACACAGTGTAGGCTTATTTGTAATTTTCATGTAGAGTAGGACCCAAAAATCCACCCCAAGT	2760
Qy	2761	CCTTTATCTATGCCACATCCTCTTATCTATCTATCTTCCAGGACACTTTTCTCTCTCTATGA	2820
Db	2761	CCTTTATCTATGCCACATCCTCTTATCTATCTTCCAGGACACTTTTCTCTCTTATGA	2820
Qy	2821	TAAGGCTCTCTCTCTCCA	2880
Db	2821	TAAGGCTCTCTCTCTCCA	2880
Qy	2881	CACAAACACACACCCCGCCACCAAGAGTGATGTAATAAGAGGATTTATTCGTAAGAGATGA	2940
Db	2881	CACAAACACACACCCCGCCACCAAGAGTGATGTAATAAGAGGATTTATTCGTAAGAGATGA	2940
Qy	2941	CTCATCTACACAGCCCGAGGGTAAAGTAAATATAGAGGGATTTATTCGTAAGAGATGA	3000
Db	2941	CTCATCTACACAGCCCGAGGGTAAAGTAAATATAGAGGGATTTATTCGTAAGAGATGA	3000
Qy	3001	TGCTTAATCTGTTTAACTCTGGGCTCAAGAGAGAAATTTCTTCTCTGTACTTATTA	3060
Db	3001	TGCTTAATCTGTTTAACTCTGGGCTCAAGAGAGAAATTTCTTCTCTGTACTTATTA	3060
Qy	3061	AGCACCCTAATATGTTTGAGCTTATATACAAAGGGTTATATATGCTTAATATAGTAAT	3120
Db	3061	AGCACCCTAATATGTTTGAGCTTATATACAAAGGGTTATATATGCTTAATATAGTAAT	3120
Qy	3121	AGTAATGKTGTTGGTACTATGGTAAATACCATAAAAAATATATCCTTTTAAAAATAAG	3180
Db	3121	AGTAATGKTGTTGGTACTATGGTAAATACCATAAAAAATATATCCTTTTAAAAATAAG	3180
Qy	3181	CTAATATATATGGATCTTTTACGATTTTATGTTTATGTTTATGTTTATGATTTTTT	3240
Db	3181	CTAATATATATGGATCTTTTACGATTTTATGTTTATGTTTATGTTTATGATTTTTT	3240
Qy	3241	AAAAGACAAATCTACCCCTGTACCCAGGCTGGAGTGCAGTGGTGCATATAGCTTTCTG	3300
Db	3241	AAAAGACAAATCTACCCCTGTACCCAGGCTGGAGTGCAGTGGTGCATATAGCTTTCTG	3300
Qy	3301	CAGTCTTCAACTCTCGGCTCAAGCAATCCTCTGCTTGGCTCCCAAGTCTTGGAT	3360
Db	3301	CAGTCTTCAACTCTCGGCTCAAGCAATCCTCTGCTTGGCTCCCAAGTCTTGGAT	3360
Qy	3361	ACAGTCATGAGCCACTGTCATCTGGCTAGGATCCATTTAGATTTAAATATGCAATTTTAA	3420
Db	3361	ACAGTCATGAGCCACTGTCATCTGGCTAGGATCCATTTAGATTTAAATATGCAATTTTAA	3420
Qy	3421	TTTTTAAATAATATGGCTAAATTTTACCTTATGTAATGTAATACGTAATAAATCTAG	3480
Db	3421	TTTTTAAATAATATGGCTAAATTTTACCTTATGTAATGTAATACGTAATAAATCTAG	3480
Qy	3481	TTTCTGCTCAAGTTTAAAGTGCTTCCCAATAGCTTCATCTAGTGAGGGAGACATT	3540
Db	3481	TTTCTGCTCAAGTTTAAAGTGCTTCCCAATAGCTTCATCTAGTGAGGGAGACATT	3540
Qy	3541	TAAAGTGAACACAGACAGCCAGGTGTGGTGCCTGATCCAGCCACTCTGGGAG	3600
Db	3541	TAAAGTGAACACAGACAGCCAGGTGTGGTGCCTGATCCAGCCACTCTGGGAG	3600
Qy	3601	GCTGAGGTGGTGATCGCTTTGAGCCCTGGAGTTCGAAGACCAGCCCTGAGCAACATGGCAA	3660
Db	3601	GCTGAGGTGGTGATCGCTTTGAGCCCTGGAGTTCGAAGACCAGCCCTGAGCAACATGGCAA	3660
Qy	3661	AACCTCTTTCTATTAACAAAATTAACCGGGCATGGTGGCATGTGCCCTGTGCTCCAGCT	3720
Db	3661	AACCTCTTTCTATTAACAAAATTAACCGGGCATGGTGGCATGTGCCCTGTGCTCCAGCT	3720
Qy	3721	ACTAGGGGCTGAGGCAGGAGAAATCTTTTGGAGCCAGGAGGTCAAGGCTGCACTGAGCAG	3780
Db	3721	ACTAGGGGCTGAGGCAGGAGAAATCTTTTGGAGCCAGGAGGTCAAGGCTGCACTGAGCAG	3780
Qy	3781	TGCTTGGGCCACTGCACCTCCAGCCCTGGGTGACAGGACACAGCTTTGCCCTCAAAAAATAA	3840

3781	Db	TGCTTGGCCCATCTGCACCTCCAGCCTCGGGTGACAGGACCGACGCTTGCCCTCAAAAAATAA	3841
3841	Qy	GAAGAAAAATTAAAAATAAATGGAACAACACTACAAAGAGCTGTGTGCTCTAGATGAGCTAC	3900
3841	Db	GAAGAAAAATTAAAAATAAATGGAACAACACTACAAAGAGCTGTGTGCTCTAGATGAGCTAC	3900
3901	Qy	TTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTACCTGCACTAC	3960
3901	Db	TTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGTACCTGCACTAC	3960
3961	Qy	ATTATTAAATATCAATTTCTCAATGTATATCCACACAAGACCTGGTGAATGTTTCAT	4020
3961	Db	ATTATTAAATATCAATTTCTCAATGTATATCCACACAAGACCTGGTGAATGTTTCAT	4020
4021	Qy	AGTACCTTTATTACAAAAACCCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAA	4080
4021	Db	AGTACCTTTATTACAAAAACCCAAAGTAGAGACTATCCAAATATCCATCAACAAGTGAA	4080
4081	Qy	CAATAAACAATAATGTGCTATATCCATGCAATGGAATACCAACCTGCGAGTACAAAGGAAG	4140
4081	Db	CAATAAACAATAATGTGCTATATCCATGCAATGGAATACCAACCTGCGAGTACAAAGGAAG	4140
4141	Qy	AAGCTACTTGGGATGAATCCCAAGTCATGACGCTAAATGAAGAGTCAGACATGAAGG	4200
4141	Db	AAGCTACTTGGGATGAATCCCAAGTCATGACGCTAAATGAAGAGTCAGACATGAAGG	4200
4201	Qy	AGGAGATAATGTATGCCATACGAAATCTAGAAAAATGAAAGTAACTTATAGTTACAGAAA	4260
4201	Db	AGGAGATAATGTATGCCATACGAAATCTAGAAAAATGAAAGTAACTTATAGTTACAGAAA	4260
4261	Qy	GCAATCAGGCGAGGCATAGAGGCTCACACCTGTAAATCCAGCACTTTGAGAGGCCACGT	4320
4261	Db	GCAATCAGGCGAGGCATAGAGGCTCACACCTGTAAATCCAGCACTTTGAGAGGCCACGT	4320
4321	Qy	GGGAAGATTCTAGAACTCAGGAGTTCAAACACAGCCTGGGCAACACAGTGAACATCCAT	4380
4321	Db	GGGAAGATTCTAGAACTCAGGAGTTCAAACACAGCCTGGGCAACACAGTGAACATCCAT	4380
4381	Qy	TCCTCCAAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTTGTCTGTGGGAGGGGAAG	4440
4381	Db	TCCTCCAAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTTGTCTGTGGGAGGGGAAG	4440
4441	Qy	GACTGCAAGAGGGAAGAAGCTCTGTGGGGTGAGGCTGGTGAATCAGGTTCTGTATCCCT	4500
4441	Db	GACTGCAAGAGGGAAGAAGCTCTGTGGGGTGAGGCTGGTGAATCAGGTTCTGTATCCCT	4500
4501	Qy	GACTGTGGTAGCAGTTTGGGGTGTATACATCCAAAAATATTCGTAGAAATATGCATCTTA	4560
4501	Db	GACTGTGGTAGCAGTTTGGGGTGTATACATCCAAAAATATTCGTAGAAATATGCATCTTA	4560
4561	Qy	AATGGGTGGAGTTTACTGTATGTAAATATACCTCAATGTAAAGAAAAATTAATGTGTAA	4620
4561	Db	AATGGGTGGAGTTTACTGTATGTAAATATACCTCAATGTAAAGAAAAATTAATGTGTAA	4620
4621	Qy	AAAAGTTTCAATTCTCTTGGCAGCAACAGTTATTCAAATTCCTGAGCCCTTTACTTCGCA	4680
4621	Db	AAAAGTTTCAATTCTCTTGGCAGCAACAGTTATTCAAATTCCTGAGCCCTTTACTTCGCA	4680
4681	Qy	AATTCCTGCACCTTCTGCCCGGTACCAATTAGTGTACAGCACTAGTCCACAAATTTGATA	4740
4681	Db	AATTCCTGCACCTTCTGCCCGGTACCAATTAGTGTACAGCACTAGTCCACAAATTTGATA	4740
4741	Qy	AATGCATTTCTGAAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTTCATATCA	4800
4741	Db	AATGCATTTCTGAAAAAGACTAGGGACAAAATCCAGGCATCACTTGTGCTTTTCATATCA	4800
4801	Qy	CCAGCTGTACAGCTTGTGTGCTGTGCTGAGCTGCAATGGGACTCTTGATTTCTTTAA	4860
4801	Db	CCAGCTGTACAGCTTGTGTGCTGTGCTGAGCTGCAATGGGACTCTTGATTTCTTTAA	4860
4861	Qy	GGAACCTTGGGTACACAGATATTTCCACAAATGCTATTCAAATTAGTGCCTTATGATATG	4920
4861	Db	GGAACCTTGGGTACACAGATATTTCCACAAATGCTATTCAAATTAGTGCCTTATGATATG	4920

QY	4921	CAAGACACTGTGCTAGGAGCCAGAAAAACAAGAGGAGGAGAAATCAGTCAATTATGTGGGA	4980
DB	4921	CAAGACACTGTGCTAGGAGCCAGAAAAACAAGAGGAGGAGAAATCAGTCAATTATGTGGGA	4980
QY	4981	ACAACATAGCAAGATATTTAGATCATTTTGACCTAGTGTAAAAAAGCAGCAGGTACAAAT	5040
DB	4981	ACAACATAGCAAGATATTTAGATCATTTTGACCTAGTGTAAAAAAGCAGCAGGTACAAAT	5040
QY	5041	CACACATGCATCAGTATAATCCAAATCATGTAATATGTCCTGTAGAAGACTAGAGG	5100
DB	5041	CACACATGCATCAGTATAATCCAAATCATGTAATATGTCCTGTAGAAGACTAGAGG	5100
QY	5101	AATAAACACACAAGATCTTTAACAGTCATTTGTCATTAGACACTAAAGTCTAATTATTATT	5160
DB	5101	AATAAACACACAAGATCTTTAACAGTCATTTGTCATTAGACACTAAAGTCTAATTATTATT	5160
QY	5161	AGACACTATGATATTTGAGATTTAAAAAATCTTTTAATATTTAAAAATTTAGAGCTCTTCT	5220
DB	5161	AGACACTATGATATTTGAGATTTAAAAAATCTTTTAATATTTAAAAATTTAGAGCTCTTCT	5220
QY	5221	ATTTTCCATAGTATTCAGGTTTGACAAATCATCAAGTATTACTCTTTCTTTTCTTTTTTT	5280
DB	5221	ATTTTCCATAGTATTCAGGTTTGACAAATCATCAAGTATTACTCTTTCTTTTCTTTTTTT	5280
QY	5281	TTTTTTTTTTTTTGAGATGGAGTTTTGGTCTTGTTCCTCATGCTGGAGTGGAAATGCATG	5340
DB	5281	TTTTTTTTTTTTTGAGATGGAGTTTTGGTCTTGTTCCTCATGCTGGAGTGGAAATGCATG	5340
QY	5341	AYCATAGCTCACTGCAACCTCCACTCTGGGTTTCAAGCAAAAGCTGTCGGCTCAGGCTCC	5400
DB	5341	AYCATAGCTCACTGCAACCTCCACTCTGGGTTTCAAGCAAAAGCTGTCGGCTCAGGCTCC	5400
QY	5401	CGGCTAGATGGGATTACAGCGGCCACACACACACACACACACACACACACACACACAC	5460
DB	5401	CGGCTAGATGGGATTACAGCGGCCACACACACACACACACACACACACACACACACAC	5460
QY	5461	GAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGAG	5514
DB	5461	GAGATGGGTTTCCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAGAG	5514
RESULT 2			
AAC55339			
ID	AAC55339	standard; DNA; 11204 BP.	
XX	AAC55339		
AC	AAC55339		
XX			
DT	05-FEB-2001	(first entry)	
XX			
DE	Human activation-induced cytidine deaminase genomic DNA	SEQ ID NO:35.	
XX			
KW	Activation-induced cytidine deaminase; AID; cytidine deaminase;		
KW	immune related disease; allergy; allergic disease; anti-allergic;		
KW	antianemic; antiasthmatic; ophthalmological; anti-HIV; dermatological;		
KW	gene therapy; B cell associated immune system disorder; food allergy;		
KW	immunodeficiency disease; immunoglobulin A deficiency disease; asthma;		
KW	IgA nephritis; gamma-globulinaemia; atopic dermatitis; allergic colitis;		
KW	drug allergy; allergic rhinitis; Kosen disease; Digorge disease; AIDS;		
KW	ataxia telangiectasia; common variable immunodeficiency disorder;		
KW	major histocompatibility class II deficiency disease;		
KW	auto immunodeficiency syndrome; IgG subclass selection disorder; ds.		
OS	Homo sapiens.		
XX			
FN	WO200058480-A1.		
XX			
PD	05-OCT-2000.		
XX			
XX	28-MAR-2000; 2000WO-JP01918.		
PF			
XX			
PR	29-MAR-1999; 99JP-0087192.		
PR	24-JUN-1999; 99JP-0178999.		

PR 27-DEC-1999; 99JP-03711382.  
XX (NISB ) JAPAN TOBACCO INC.  
PA (HONJ/) HONJO T.  
XX  
XX  
PI Honjo T, Muramatsu M;  
XX WPI; 2000-611715/58.  
XX  
XX  
XX Nucleic acid encoding activation induced cytidine deaminase, useful as  
PT a target for drug development for immune-related diseases including  
XX allergies -  
XX  
XX  
PS Claim 17; Page 163-170; 174pp; Japanese.  
XX  
XX The present invention describes an activation-induced cytidine deaminase  
CC (AID). AID structurally relates to an RNA editing enzyme APOBEC-1 and  
CC has cytidine activity similar to APOBEC-1. AID has anti-allergic,  
CC antianemic, antidiabetic, ophthalmological, anti-HIV and  
CC dermatologic activities, and can be used in gene therapy. AID  
CC polynucleotides are useful in methods for identifying drugs for the  
CC treatment of B cell associated immune system disorders, immunodeficiency  
CC diseases and allergies, such as immunoglobulin A (IgA) deficiency  
CC disease, IgA nephritis, gamma-globulinemia, atopic dermatitis, allergic  
CC colitis, asthma, food allergy, drug allergy, allergic rhinitis, Rosen  
CC disease, DiGeorge disease, ataxia telangiectasia, common variable  
CC immunodeficiency disorder, MHC (major histocompatibility class) class  
CC II deficiency disease, AIDS (auto immunodeficiency syndrome), elevated  
CC IgE disorder, and IgG subclass selection disorder. The DNA sequences  
CC encoding AID may be used for gene therapy and the antibodies to the AID  
CC protein may be used for diagnosis and treatment of these disorders. The  
CC present sequence represents a genomic DNA sequence of human AID.  
XX  
XX Sequence 11204 BP; 3305 A; 2273 C; 2373 G; 3253 T; 0 other;

Query Match 89.3%; Score 4922.8; DB 21; Length 11204;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 4921; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

591 AGGTTACAGAGACTGTGGAAATATGGGGAAATAGAGGCTATCTGAGGCTCTTCAACAC 650  
1 AGGTTACAGAGACTGTGGAAATATGGGGAAATAGAGGCTATCTGAGGCTCTTCAACAC 60  
651 AATAACCCAAAGAGCTATTTAAATGCTCTTAAAGTATTACATAAATATTACTATTCTC 710  
61 AATAACCCAAAGAGCTATTTAAATGCTCTTAAAGTATTACATAAATATTACTATTCTC 120  
711 ATTGCTCTTTATTTGTTATCATGATTATTAATGAAGTGTCTACTGTTACTGCTCC 770  
121 ATTGCTCTTTATTTGTTATCATGATTATTAATGAAGTGTCTACTGTTACTGCTCC 180  
771 TGATCTTTGCTAGCTATGGAGCATGGACTGGCTTTTATAGGACGACGCCCAAGGAACC 830  
181 TGATCTTTGCTAGCTATGGAGCATGGACTGGCTTTTATAGGACGACGCCCAAGGAACC 240  
831 TAAACATTAAGCAGAGCTGCCCTCAATGGTTTAACTGTTGAGCTCTGCTATGACAGC 890  
241 TAAACATTAAGCAGAGCTGCCCTCAATGGTTTAACTGTTGAGCTCTGCTATGACAGC 300  
891 CCCACCCACCCTCTTCACCTGGATCCAAATCAGGAGCAAGCCGTTGGGGTACCTGGTGG 950  
301 CCCACCCACCCTCTTCACCTGGATCCAAATCAGGAGCAAGCCGTTGGGGTACCTGGTGG 360  
951 GGGTGATGCTGTACGGGGAGGAGCCCAAGGGCAAGCTCAAAATTTGAATGTGAAGGCC 1010  
361 GGGTGATGCTGTACGGGGAGGAGCCCAAGGGCAAGCTCAAAATTTGAATGTGAAGGCC 420  
1011 AATGCATGTCAGACTGAGACAGAGAACCATCATTAATTTGAAGTGAGATTTTCTGGCCT 1070  
421 AATGCATGTCAGACTGAGACAGAGAACCATCATTAATTTGAAGTGAGATTTTCTGGCCT 480  
1071 GAGACTTGCAGGGAGGCAAGAGACACTCTGGACACCACCTATGGACAGGTTAAGAGGCAG 1130

481 GAGACTTGCAGGAGGCAAGAGACACTCTGGACACCACCTATGGACAGGTAAAGAGCAG 540  
1131 TCCTTCGTCGGTGATTCGACTGCGCTTCCCTCTCAGAGCAAACTCGAGTAATGAGACTGG 1190  
541 TCCTTCGTCGGTGATTCGACTGCGCTTCCCTCTCAGAGCAAACTCGAGTAATGAGACTGG 600  
1191 TAGCTATCCCTTTCTCTCATGTAACCTGCTGACTGATAAGATCAGCTTGTATCAATATGCA 1250  
601 TAGCTATCCCTTTCTCTCATGTAACCTGCTGACTGATAAGATCAGCTTGTATCAATATGCA 660  
1251 TATATATTTTTCATGTCCTCCTTTCTTCTTATTCAGATCTTATACGCTGTACGCCAA 1310  
661 TATATATTTTTCATGTCCTCCTTTCTTCTTATTCAGATCTTATACGCTGTACGCCAA 720  
1311 TTCTTCTGTTTCAGACTCTCTTGTATTTCCCTCTTTCATGTGGCAAAAGAGTACTG 1370  
721 TTCTTCTGTTTCAGACTCTCTTGTATTTCCCTCTTTCATGTGGCAAAAGAGTACTG 780  
1371 CGTACAATGTACTGCTCTCTGAGATTTGTACCATGTTGTAAACTAAATTTATGTTAAT 1430  
781 CGTACAATGTACTGCTCTCTGAGATTTGTACCATGTTGTAAACTAAATTTATGTTAAT 840  
1431 AATATTAACATAGCAAAATCTTTAGAGACTCAAAATCATGAAAAGTAAATAGCAGTACTGTA 1490  
841 AATATTAACATAGCAAAATCTTTAGAGACTCAAAATCATGAAAAGTAAATAGCAGTACTGTA 900  
1491 CTAAAAACGCTAGTGTCTAAATTTTCGTAATTAATTTGTAAATATTTCAACAGTAAACAACT 1550  
901 CTAAAAACGCTAGTGTCTAAATTTTCGTAATTAATTTGTAAATATTTCAACAGTAAACAACT 960  
1551 TGAAGACACACTTTCTCTAGGAGGCGTTACTGAAATAAATTTAGCTATAGTAAAGAAATTT 1610  
961 TGAAGACACACTTTCTCTAGGAGGCGTTACTGAAATAAATTTAGCTATAGTAAAGAAATTT 1020  
1611 GTAAATTTAGAAATGCCAAGCATCTTAAATTAATTTGCTTGAAGTCACTATGATGTGTC 1670  
1021 GTAAATTTAGAAATGCCAAGCATCTTAAATTAATTTGCTTGAAGTCACTATGATGTGTC 1080  
1671 CATTATAAGGAGACAAATTCATTCAAGCAAGTATTATTAATTTAAAGGCCCAATTTGTAG 1730  
1081 CATTATAAGGAGACAAATTCATTCAAGCAAGTATTATTAATTTAAAGGCCCAATTTGTAG 1140  
1731 GCAGTTAATGGCACTTTTACTATTAACATAATCTTCCATTTGTTTCCAGACGCTTAACT 1790  
1141 GCAGTTAATGGCACTTTTACTATTAACATAATCTTCCATTTGTTTCCAGACGCTTAACT 1200  
1791 TACCTCTTAGTGGAATTTGGTTAAGGTCCTCAATGCTCTTATGTCAGTCTTTCAT 1850  
1201 TACCTCTTAGTGGAATTTGGTTAAGGTCCTCAATGCTCTTATGTCAGTCTTTCAT 1260  
1851 AGGTTATGTGATAGAACTTATCTTATCTTACATTTATGATTTACTATGGATGATGAGA 1910  
1261 AGGTTATGTGATAGAACTTATCTTATCTTACATTTATGATTTACTATGGATGATGAGA 1320  
1911 ATACACCTAATCCTTATACCTTACCTCAATTTAACTCCTTTTATAAAGAACTTACATTTAC 1970  
1321 ATACACCTAATCCTTATACCTTACCTCAATTTAACTCCTTTTATAAAGAACTTACATTTAC 1380  
1971 AGAATAAGATTTTTTAAAAATATATTTTTTGTAGACAGAGGCTTAGCCAGCCGAG 2030  
1381 AGAATAAGATTTTTTAAAAATATATTTTTTGTAGACAGAGGCTTAGCCAGCCGAG 1440  
2031 GCTGGTCTCTAAGTCTGGCCCAAGCGATCCTCCTGCTGGCCCTCTCTAAAGTGTGGAA 2090  
1441 GCTGGTCTCTAAGTCTGGCCCAAGCGATCCTCCTGCTGGCCCTCTCTAAAGTGTGGAA 1500  
2091 TTATPAGACATGAGCCATCACAATCAATATACAGAAATAAAGATTTTTTAAAGGAGATTTAA 2150  
1501 TTATPAGACATGAGCCATCACAATCAATATACAGAAATAAAGATTTTTTAAAGGAGATTTAA 1560  
2151 TGTTCCTCAGAAAAATTTCTTTCAGGTGAGCAATGTCAAATGTCCTCCTCAGTTTACACTG 2210  
1561 TGTTCCTCAGAAAAATTTCTTTCAGGTGAGCAATGTCAAATGTCCTCCTCAGTTTACACTG 1620

QY	2211	AGATTTTGA	AAACAAAGTCTGAGCTATAGGTCTCTTGTGAAGGGTCCATTGGAAATAC	TGTGT	2270
DB	1621	AGATTTTGA	AAACAAAGTCTGAGCTATAGGTCTCTTGTGAAGGGTCCATTGGAAATAC	TGTGT	1680
QY	2271	TCAAGTAA	ATAATGCAAGCAAGGTAAATCAGCAGTTGAAATTCAGAGAAAGACACAAA	2330	
DB	1681	TCAAGTAA	ATAATGCAAGCAAGGTAAATCAGCAGTTGAAATTCAGAGAAAGACACAAA	1740	
QY	2331	AGGAGAAA	AGATGAAATTCACAGCACAGAGGGAATATATTATCATTAAGAGGACAG	2390	
DB	1741	AGGAGAAA	AGATGAAATTCACAGCACAGAGGGAATATATTATCATTAAGAGGACAG	1800	
QY	2391	TATCTGTAG	AGCTCATTTAGTGATGGCAAAATGACTTGGTCAGGATATTTTTAACCCGCT	2450	
DB	1801	TATCTGTAG	AGCTCATTTAGTGATGGCAAAATGACTTGGTCAGGATATTTTTAACCCGCT	1860	
QY	2451	TGTTTCTCG	TTGTCACGGCTGGGATGCAGCTAGGGTCTGCCTCACGGGAGCACAGCTGT	2510	
DB	1861	TGTTTCTCG	TTGTCACGGCTGGGATGCAGCTAGGGTCTGCCTCACGGGAGCACAGCTGT	1920	
QY	2511	CCAGAGCAG	CTGTGAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCTCTTCC	2570	
DB	1921	CCAGAGCAG	CTGTGAGCCTGCAAGCCTGAAACACTCCCTCGGTAAAGTCTCTTCC	1980	
QY	2571	GGACAGAA	ATGACGAGAACAGGGAGCTGGAAACAGGCCCTTAACCCAGAGAGGAAGTAA	2630	
DB	1981	GGACAGAA	ATGACGAGAACAGGGAGCTGGAAACAGGCCCTTAACCCAGAGAGGAAGTAA	2040	
QY	2631	TGGATCAAC	AAAGTTAACTAGCAGGTACAGATCAGGCAATTCATTTCACTCTGACTGGTA	2690	
DB	2041	TGGATCAAC	AAAGTTAACTAGCAGGTACAGATCAGGCAATTCATTTCACTCTGACTGGTA	2100	
QY	2691	ACATGTGAC	AGAAAACAGTGTAGGCTTATTTGATTTTCATGTAGAGTAGGACCCAAAATC	2750	
DB	2101	ACATGTGAC	AGAAAACAGTGTAGGCTTATTTGATTTTCATGTAGAGTAGGACCCAAAATC	2160	
QY	2751	CACCCAAG	TCCTTATCTATGCCACATCCCTTTATCTATCTTCCAGGACACTTTTTC	2810	
DB	2161	CACCCAAG	TCCTTATCTATGCCACATCCCTTTATCTATCTTCCAGGACACTTTTTC	2220	
QY	2811	TTCCCTTAT	GATAAGGCCTCTCTCTCTCCACACACACACACACACACACACACACACA	2870	
DB	2221	TTCCCTTAT	GATAAGGCCTCTCTCTCTCCACACACACACACACACACACACACACACA	2280	
QY	2871	CACACACAC	ACAAAACACACACCCGCCCAACCAAGGTGCATGTAAAAAGATGTAGATTC	2930	
DB	2281	CACACACAC	ACAAAACACACACCCGCCCAACCAAGGTGCATGTAAAAAGATGTAGATTC	2340	
QY	2931	CTCTGCCCT	TTCTCATCTACAGCCCAAGGAGGTAAAGTTAATAAGAGGGATTTATGG	2990	
DB	2341	CTCTGCCCT	TTCTCATCTACAGCCCAAGGAGGTAAAGTTAATAAGAGGGATTTATGG	2400	
QY	2991	TAAGAGATG	ATGCTTAATCTGTTTAAACACTGGGCCCTCAAAGAGAGAATTTCTTTCTCT	3050	
DB	2401	TAAGAGATG	ATGCTTAATCTGTTTAAACACTGGGCCCTCAAAGAGAGAATTTCTTTCTCT	2460	
QY	3051	GTACTTATTA	AGCACCTTATGTGTTGAGCTTATATATACAAAGGGTATTATATGCTA	3110	
DB	2461	GTACTTATTA	AGCACCTTATGTGTTGAGCTTATATATACAAAGGGTATTATATGCTA	2520	
QY	3111	ATATAGTAA	TAGTAAATGKTGGTGGTACTATGGTAATACCATAAAAATTAATTCCTTT	3170	
DB	2521	ATATAGTAA	TAGTAAATGKTGGTGGTACTATGGTAATACCATAAAAATTAATTCCTTT	2580	
QY	3171	TAAAAATAA	AGCTAAATTAATTTGGATCTTTTTTAGTATTCATTTTATGTTTTTATGCTT	3230	
DB	2581	TAAAAATAA	AGCTAAATTAATTTGGATCTTTTTTAGTATTCATTTTATGTTTTTATGCTT	2640	
QY	3231	TTGATTTTT	AAAAAGACAACTCACCCCTGTATACCCAGGCTGGAGTGCAGTGGTCAATCA	3290	
DB	2641	TTGATTTTT	AAAAAGACAACTCACCCCTGTATACCCAGGCTGGAGTGCAGTGGTCAATCA	2700	

QY	3291	TAGCTTCTTCGAGCTCTTGAACTCCTGGGCTCAAGCAATCCTCCTGCTTGGCCTCCCAA	3350
DB	2701		
		TAGCTTCTTCGAGCTCTTGAACTCCTGGGCTCAAGCAATCCTCCTGCTTGGCCTCCCAA	2760
QY	3351	GTGTTGGGATACAGTCAATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTTAAATAT	3410
DB	2761		
		GTGTTGGGATACAGTCAATGAGCCACTGCATCTGGCCTAGGATCCATTTAGATTTAAATAT	2820
QY	3411	GCATTTTAAATTTTAAATTAATATGGCTAAATTTTACCTTTATGTAAATGTATCTACTGGTA	3470
DB	2821		
		GCATTTTAAATTTTAAATTAATATGGCTAAATTTTACCTTTATGTAAATGTATCTACTGGTA	2880
QY	3471	ATAAATCTAGTTTTCCTGCCCTAAAGTGTAAAGTGCTTTTCCAATTAAGCTTCATGTACGTGAG	3530
DB	2881		
		ATAAATCTAGTTTTCCTGCCCTAAAGTGTAAAGTGCTTTTCCAATTAAGCTTCATGTACGTGAG	2940
QY	3531	GGGAGACATTTAAAGTGAACACAGACGCCAGGTGTGGTGCTCACGCCCTGTATCCGAGC	3590
DB	2941		
		GGGAGACATTTAAAGTGAACACAGACGCCAGGTGTGGTGCTCACGCCCTGTATCCGAGC	3000
QY	3591	ACTCTGGGAGGCTCAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACACGCCCTGAGC	3650
DB	3001		
		ACTCTGGGAGGCTCAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACACGCCCTGAGC	3060
QY	3651	AACATGCCAAACCCCTCTTTCTATAACAAAATTTAGCCGGCATGTGGCATGTGCCCTGT	3710
DB	3061		
		AACATGCCAAACCCCTCTTTCTATAACAAAATTTAGCCGGCATGTGGCATGTGCCCTGT	3120
QY	3711	GGTCCCAGCTACTAGGGGGCTGAGGCAGGAGAATCTTTGGAGCCAGGAGGTCAAGGCTG	3770
DB	3121		
		GGTCCCAGCTACTAGGGGGCTGAGGCAGGAGAATCTTTGGAGCCAGGAGGTCAAGGCTG	3180
QY	3771	CACTGACAGCTGTTCGCCACATCGACTCCAGCCTGGGTGACAGGACCAGACCTTGCCTC	3830
DB	3181		
		CACTGACAGCTGTTCGCCACATCGACTCCAGCCTGGGTGACAGGACCAGACCTTGCCTC	3240
QY	3831	AAAAAATAAGAGAAAAATTTAAAAATPAAATGGAACAACACTACAAAGAGCTGTGTCCCTA	3890
DB	3241		
		AAAAAATAAGAGAAAAATTTAAAAATPAAATGGAACAACACTACAAAGAGCTGTGTCCCTA	3300
QY	3891	GATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTTAACTTTTAAAGTCAGGCTCTGTCA	3950
DB	3301		
		GATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTTAACTTTTAAAGTCAGGCTCTGTCA	3360
QY	3951	CCTGCACCTACATTTAAATATCAATTCCTCAATGTATATCCACACAAAGACTGGTACGT	4010
DB	3361		
		CCTGCACCTACATTTAAATATCAATTCCTCAATGTATATCCACACAAAGACTGGTACGT	3420
QY	4011	GAATGTTTCATAGTACCTTTATTCACAAACCCCAAGTAGAGACTATCCAAATATCCATC	4070
DB	3421		
		GAATGTTTCATAGTACCTTTATTCACAAACCCCAAGTAGAGACTATCCAAATATCCATC	3480
QY	4071	AACAAGTGAACAAATAACAAAATGTGCTATATCCATGCAATGGAATACCACCTCGAGT	4130
DB	3481		
		AACAAGTGAACAAATAACAAAATGTGCTATATCCATGCAATGGAATACCACCTCGAGT	3540
QY	4131	ACAAAGGAAGAGCTACTTCGGGATGAATCCCAAAGTCATGACGCTAAATGAAAGAGTCA	4190
DB	3541		
		ACAAAGGAAGAGCTACTTCGGGATGAATCCCAAAGTCATGACGCTAAATGAAAGAGTCA	3600
QY	4191	GACATGAAGGAGGAGATAATGTATGCCATACGAAATTTAGAAAAATGAAAGTAACTTATA	4250
DB	3601		
		GACATGAAGGAGGAGATAATGTATGCCATACGAAATTTAGAAAAATGAAAGTAACTTATA	3660
QY	4251	GTTTACAGAAAGCAATCAGGGCAGGCATAGAGGCTCACACCTGTAAATCCAGACACTTTGA	4310
DB	3661		
		GTTTACAGAAAGCAATCAGGGCAGGCATAGAGGCTCACACCTGTAAATCCAGACACTTTGA	3720
QY	4311	GAGCCACGCTGGGAGATTCCTAGAAGCTCAGGAGTTCAAGACCAGCCTGGGCAACACGT	4370
DB	3721		
		GAGCCACGCTGGGAGATTCCTAGAAGCTCAGGAGTTCAAGACCAGCCTGGGCAACACGT	3780
QY	4371	GAAACTCCATTTCCCAAAAAATTTGGGAAAAAAGAAAGCAAAATACGTGGTGTGCTGTG	4430









Qy	3488	CCTAAGTTTAAAGTGTCTTCCAAATAGCTTCATGTCAGTGGGGGAGACATTTTAAAGTG	3547	Qy	4548	ATTATGCATCTTAAATGGGTGGAGTTTACTGTATGTAAATATACCTCAANTGTAAGAAA	4607
Db	19843	AAATAATGAATAGGAAAGCTTTCTTAAGCCCTACTTTCCAAATCGGAAACCAAAACAA	19784	Db	18794	AAAAAGAAAAAAGAAAAATCTAAACTATATCCAAAATATTTAAAGTTTAACTCT	18735
Qy	3548	AAACAGACA-----GCCAGTGTGGTGGCTCACGCCCTGTAAATCCAGACACTCTGG	3598	Qy	4608	ATAATGTGTAGAAAAAGTTTCAATTCCTTCCAGCAACCTTATTTCAAATTCCTGAGC	4667
Db	19783	AAAGCCACACAGGCTTAGGCCAGATGTGGTGGCTCATGCCGTGAATCCAGACCTTTGGG	19724	Db	18734	TTAGACTTTTGGCCACTGTCTATATTTAGATTTAGAAATAGAAAAAGCAATTTTTT	18675
Qy	3599	AGGCTGAGGTGGTGGATCGCTTGAAGCCCTGGAGTTCAAGACCACCTTGAGCAACATGGC	3658	Qy	4668	CGTTTACTTCGCAAAATCTCTGCACCTTCCTGCCCCCTGACCATTTAGTGTGACGACCTAGCTC	4727
Db	19723	AGGCTTAAGCGAGTGGATCACCTTGAGATCAGAGTTGAGGACCGACCTGGCCAACTGGT	19664	Db	18674	TCAAATTTGGAAAAATTTGATAGCCCTTTTAACAGAGTAATTTACACTTTGTAGACATTTTC	18615
Qy	3659	AAAAACCTGTCTCT-----ATACAAAAAATTTAGCCGGCATGTGGCATGTGCTGG	3712	Qy	4728	CACAAATTTGGATAAATGCAATTTCTGGAAAAAGACTAGGGACAAAA-----TCCAGGCA	4779
Db	19663	GAGACCCCATCTCTACTAAACTACAAAAATTTAGTAGAGTGGTGTGGGTGGCTGTAA	19604	Db	18614	CTAAAGAAATAATCAGAGATTGCATAAAGATACATTTATAAGAATCGTAATACCAAAA	18555
Qy	3713	TCCAGCTACTAGGG--GGCTGAGCGAGGAGAACTTTTGGAGCCCGAGGAGGTCAAGCTGC	3771	Qy	4780	TCACTTGTGCTTTCATATCAACACGCTGTACAGCTTTGTGCTGTCTGCAGCTGCAAT	4839
Db	19603	TCCAGCTACTTTGGGAGTTGGGCGAGGAGAAATGCTTGAAGCCAGGAGTCAGAGTTGC	19544	Db	18554	AAATTGAAATTCACAAAAATCTAACTACCTGGGAATTTGGTTACATAAATATGCCACATTCATAT	18495
Qy	3772	ACTGAGCAGTGTTCGGCCACTGCACCTCCAGCTGGGTGACAGACGACCTTGCCTCA	3831	Qy	4840	GGGACTCTTGATTTCTTTAAGGAAACT--TGGGTTACCAGAGTATTTTCCACAAATGCTAT	4898
Db	19543	AATGAACCGAGTCCGCCATTTGCCATCCACCTGGGTGACAGGTGAGACTCTGTCTCT	19484	Db	18494	GATGAAAAAGTATGCACTACTAAAATTTATGACATGCAAGAATATTGACATGGAAAAAGTG	18435
Qy	3832	AAAAAATAAGAAAGAAAAATTTAAAAATAAATGGAACAACTACAAAGAGCTGTGTCCCTAG	3891	Qy	4899	TCAAATTTAGTGTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAAACAAGAGAGG	4958
Db	19483	AAATAAATAAGC---TTCAATCATAGATAAATAAAGAAAAATAAATACTGGA	19428	Db	18434	TCACACTATCTGTACATACAGGGTACAGAAACATATATAGTATGATACCTTTTGAAT	18375
Qy	3892	ATGAGCTACTTATAGTGTATATTTTGGTATTTTAACTTTTAAAGTCAAGGTCTGTCTAC	3951	Qy	4959	AGAAATCAGTCAATTTGTCGGAACAACATAGCAAGATATTTAGATCATTTTGTACTAGTTA	5018
Db	19427	AAAGGATATGATTAACAGATAGGACTGAGGATTTTCTTACTTATAAAAAAGCTATTTAT	19368	Db	18374	AGAAATTCAGATGACCGTGTATGTGTGTGATATGTCGACATATGTCGCAACATATACCASTTA	18315
Qy	3952	CTGCACTACATTTTAAATATCAATTCATGATGATATCCACAAAGACTGGTACGTG	4011	Qy	5019	AAAAAGCAGCAGAGTACAAAAATCACACATGCAATCAGTATTAATCCAAATCATGTAATAAT	5078
Db	19367	ATACCAATTAATAGGAAGATGTGAACACAGCAAAAATATCAAGGACATA-----AATAG	19313	Db	18314	TCCTTACATAGTGGGTTATGGGTAATAATAAATGTTTATCTTTGGGCTCTTCTATCTTT	18255
Qy	4012	AATGTTCTATGATACCTTTTATTCACAAAACCCAAAGTAGAGACTATCCAAATATCCATCA	4071	Qy	5079	GTGCTGTAGAAGACTAGAGGAATAACACAAGAATCTTAACAGTCAATTGATGATAGAC	5138
Db	19312	AAATTCATGACAGTGGTAAAAAAATTCAA-----ACTCACTAATAATCAA	19266	Db	18254	-TCCAAAGTTTCTTAAAGACCTGTATCTTTTACATTTGGGGGTCAAGGCTGTGTTTT	18196
Qy	4072	ACAAGTGAACAAATAACAAAAATGTGCTATATCCATGCAATGGAATPACCCCTGGAGTA	4131	Qy	5139	ACTAAGCTAAATTTATTATTATAGACACTATGATATTGAGATTTTAAAAAATCTTTAATA	5198
Db	19265	AGGAGTACAAATTTGAAATGAGACCTTTTCTGTCTAGAAAATTTAATAATGGAATAATCCA	19206	Db	18195	ATTTTATTTATTTTATTTTACAGCATGGCCACTATAATAATGATATATATGCTGCAA	18136
Qy	4132	CAAGGAAGAAGCTACTTGGGATGAATCCCAAAGTCATGAGCTAAATGAAAAGAGTCAG	4191	Qy	5199	TTTTAAAAATTTAGAGCTCTCTATTTTTCATAGTATTCAAAGTTTGGACA-----	5247
Db	19205	GAATTTGGGATATAGGAAGTAACACTCATATACTGATCATGGGAATATAAGTGGCAG	19146	Db	18135	TTTGCTAAGAGAGATAAACCTGTGTTTTTCTTTTGCATCTCTTTTGTATTTTAAATTTT	18076
Qy	4192	ACATGAAGGAGAGATAATGTATGCCATACGAAATTTCTAGAAAATGAAAGTAACCTATAG	4251	Qy	5248	-----ATGATCAAGTATTACTCTCTTTTCTTTT	5274
Db	19145	AAATCTGCAAGACACTTGAATATCTATCAAAAGCTTTAAAGAGAA-----GCAT	19095	Db	18075	TGTGGGTATATAATAGTGTGCTTTTATTTTATTTTATTTTATTTTATTTTATTTTAT	18016
Qy	4252	TTACAGAAAGCAAAATCAGGCGAGGCATAGAGCTCACACCTGTATCCCGACACTTTGAG	4311	Qy	5275	TTTTTTTTTTTTTTTTTTTTCAGATGGAGTTTTCGTGTTGTCCTGATGCTGGAGTGGAAAT	5334
Db	19094	TTCTAAAGAAAAACCTTAGGCTGGGTGGCGATGCTCACACCTGTATATCCCGACACTTTGG	19035	Db	18015	TTATTTATTTATTTTATTTTTCAGATCGAGTCTCGCTGTGTTGCCAGTTTGTAGTGCAGT	17956
Qy	4312	AGGCCACGTGG--GAAGATGCTAGAACTCAGGAGTTCAAGACAGCCTGGGCAACACAGT	4370	Qy	5335	GGCATGAYCATAGCTCACTCAACCTCCACCTCCCTGGGTTCAAGCAAAAGCTGTGCGCTCA	5394
Db	19034	AGGCCAAGCAAGTGTATCTTTGAGCCAGGAGTTTGAGACCAGGCTGGCTTAACATGGC	18975	Db	17955	GGCGCAATCTCAGCTCACTGCAACCTCCACCTCCCTGGTTCAGAGCAATTCCTCGCTCA	17896
Qy	4371	GAACCTCCATTTCCACAAAAATGGGAAAAAAGAAAGCAATTCAGTGGTTGTCGTGG	4430	Qy	5395	GCCTCCGGGTAGATGGGATACAGGGCCCCACACACTCGGCTTAATGTTTGTATTTT	5454
Db	18974	AAAACCGTCTCTCTACTAAATAATACAAAAATTAGCTGGGCTGGTGGCACACACCTGTAG	18915	Db	17895	GCCTCCCTGTTAGCTGGGACTCAGGCACGACACACACACAGCTAACATTTGTATTT	17836
Qy	4431	GGAGGGGAA---GGACTGCAAGAGAGGAAGAGCTCTGTGGGGTGAGGGTGGTGAATCA	4487	Qy	5455	TTAGTAGAGATGGGGTTTCCACCATGTTGGCAGGCTGGTCTCAAACTCCGACCTCA	5511
Db	18914	TTCCAGCTACTTTGGGAGGTGAGCCAGGAGATTTGCTAGAACCCAGGAGTTGGAGTTGC	18855	Db	17835	TTAGTAGAGATGGGGTTTCCACCATGTTGGCAGGCTGGTCTGGAACTCCCTGACCTCA	17779
Qy	4488	GGTTCTGTATCCTGACTGTGGTAGCAGTTTGGGCTGTTTACATCCAAAAAATATTCGTAGA	4547				
Db	18854	AATGAGCTGAGATCACACCACCTGCACCTCCAGCTTAGGTGACAGGGAGTCTCAGTCTCGAT	18795				

RESULT 5  
ABK83567/C  
ID ABK83567 standard; cdna; 112460 BP.  
XX



XX 09-AUG-2001.  
PD 17-JAN-2001; 2001WO-US01354.  
XX  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0235837.  
PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0238802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0256719.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Barash SC, Ruben SM;

XX  
XX  
XX  
XX

WPI; 2001-483426/52.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides, useful for preventing, diagnosing and/or treating cancers and metastasis -

Disclosure; SEQ ID NO 34974; 3071pp + Sequence Listing; English.

AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I) amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic activity, and can be used in gene therapy and vaccine production. (I) proteins and polynucleotides may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate (I) expression. For example, they may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of (I) by expressing inactive proteins or to supplement the patient's own production of (I). Additionally, (I) polynucleotides may be used to produce the secreted (I), by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) proteins and polynucleotides may be used to prevent, diagnose and treat immune/haematopoietic-related diseases, especially cancers and cancer metastases of haematopoietic-derived cells. AAK64703 to AAK87694 represent human immune/haematopoietic antigen genomic sequences from the present invention. AAK54942 to AAK54950 and AAK82169 represent sequences used in the exemplification of the present invention.

Sequence 4126 BP; 1237 A; 760 C; 851 G; 1278 T; 0 other;

Query Match	4.1%	Score	227.6	DB	22	Length	4126
Best Local Similarity	65.2%	Pred. No.	5.8e-39				
Matches	421	Conservative	0	Mismatches	204	Indels	21
Gaps	5						
QY	3213	TTTATGTTTTTTATGTTTTTTGATTTTTTAAAGACAAATCTCACCGTCTGTACCCAGGCTGG	3272				
Db	1685	TATAATTTTTTTTTTTTTTTTTTTTTTTTGGAGCGGAGTCTCGCTCTGTCCACGAGGCTGA	1626				
QY	3273	AGTGCAGTGGTGCATATAGCTTTCTGCAGTCTTGAACCTCTGGGCTCAAGCAATCCCTC	3332				
Db	1625	AGTGCAGTGGCACCATCAAGCACTCAGTGCAGCCTTCACCTCCTGGGCTCAAGCAATCCCTC	1566				
QY	3333	CTGCGTTGGCTCCCAAGGTGTTGGGATACAGTCAATGAGCCACTGCATCTGGGCTAGGAT	3392				
Db	1565	CTACCTCAGCCTCCCAAGTAGCTGG-----AAACAGGCGATGCACCCCTGTGCTCAACT	1511				
QY	3393	CCATTTAGATTAAAAATATGCATTTTAAATTTTAAAAATAATATGGCTAAATTTTACCTTAT	3452				
Db	1510	CTTTTTTTTTTTTTTTTTTTTTTTTTTAAATAGGCACAGGATCTCCCTATGTGCTCCAGCT	1451				
QY	3453	GTAATGTGATACTGTAATAAATCTAGTTTGTCTGCTTAAAGTTTAAAGTGTCTTCCAAT	3512				
Db	1450	GGTCTTGAACTACTGG-----GCCTAAGTGAATCTCCACACTTGGCTCCCAAAAGTGT	1398				
QY	3513	AAGCTTCATGCTGAGGGGAGACATTTTAAAGTGAACACAGACGCGAGTGTGGTGGCT	3572				
Db	1397	GAGCCACTGCACCTGCCTATATTTTATTTAATAAAGTTTCAAGTGGCGTGGTGGCT	1338				
QY	3573	CAGCGCTGTAATCCAGCACTCTGGAGGCTGAGTGGTGGTGGATCGTGTAGCCCTGGAG	3632				
Db	1337	CATGCGCTGTAATCCAGCAATTTGGGAGGCTGAGCGGGCAGCATCAC--GAGGTCAAGG	1280				
QY	3633	TTCAAGACCAAGCTCTGAGCAACATGGCAAAACCTGTTTCT-----ATACAAAAATTAG	3686				
Db	1279	TTGAGGCCAGCCTGACCAACATGGTCAAAACCTGTCTCTACTTAAATAACAAAATTAG	1220				
QY	3687	CCGGGCAATGTCATGTGTCCTGTGTCCAGCTACT-AGGGGGCTGAGCGAGGAAATC	3745				
Db	1219	CCAGACATGTTGTGGCTGCCTCTAATCCAGCTACTCAGGAGGCTGAGCGAGGAAATC	1160				
QY	3746	TTTGGAGCCACAGGTTCAAGGCTGCACCTGACGAGTGTGTCGCCCACTGCATCCAGCCT	3805				
Db	1159	ACTTGAAACCCAGAGGTGGAGGTGTAGTAAAGCTTGAGCTTGAGCCACTGCATCCAGCCT	1100				
QY	3806	GGGTGACGAGGACAGACCTTGGCTCAAAAAAATAAGAAAGAAAAAT	3851				



PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249266.

PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-502630/55.  
XX  
XX Polynucleotides encoding digestive system antigens, useful for  
PT diagnosing, treating, preventing and/or prognosing disorders of the  
PT digestive system, particularly cancer and cancer metastases -  
XX  
XX Disclosure; SEQ ID NO 3037; 986pp; English.  
XX  
XX The present invention provides the protein and coding sequences of a  
CC number of human digestive system antigens. These can be used in the  
CC diagnosis, treatment and prevention of digestive system disorders,  
CC including cancer, Meckel's diverticulum, bacterial or parasitic  
CC infections, appendicitis, Hirschsprung's disease, chronic colitis or  
CC ulcerative colitis. The present sequence is a genomic DNA fragment  
XX encoding a digestive system antigen of the invention.  
XX  
SQ Sequence 4126 BP; 1237 A; 760 C; 851 G; 1278 T; 0 other;  
Query Match 4.1%; Score 227.6; DB 22; Length 4126;  
Best Local Similarity 65.2%; Pred. No. 5.8e-29;  
Matches 421; Conservative 0; Mismatches 204; Indels 21; Gaps 5;  
QY 3213 TTTATGTTTTTTATGTTTTTTGATTTTTTAAAGACAAATCTACCCCTGTACCCAGGCTGG 3272  
DB 1685 TATAATTTTTTTTTTTTTTTTTTTTGTAGACGGAGTCTCGCTCTGTCAACCCAGGCTGA 1626  
QY 3273 AGTGCAGTGGTGAATCATAGCTTTCTGCAGCTTGAACCTCTGGGCTCAAGCAATCCTC 3332  
DB 1625 AGTGCAGTGGGACCATCAGAACTCAGTGCAGGCTTCACTCTGGGCTCAAGCAATCCTC 1566  
QY 3333 CTGCCTTGGCTCCCAAGTGTGGGATACAGCATGAGCCACTGTCATCTGGCCTAGGAT 3392  
DB 1565 CTACCTCAGCCTCCCGAGTAGCTGG-----AAACACAGGCGATGCACCCCTGTGCTCAACT 1511  
QY 3393 CCATTTAGATTAAATAATGATTTTAAATTTTAAATAATATGGCTAATTTTACCTTAT 3452  
DB 1510 CTTTTTTTTTTTTTTTTTTTTTTTTAAATTAAGGACAGGATCTCCCTATGTTGCTCCAGCT 1451  
QY 3453 GTAATGTGTACTAGTAAATAATCTAGTTTGTGCTTAAGTTTAAAGTCTTTCCAAT 3512  
DB 1450 GGTCTTGAACACTAGG-----GCTCAAGTGATTTCTCCACCTTGGCCTCCCAAGTGT 1398  
QY 3513 AAGCTTCATGTAGCTAGGGGAGACATTTAAAGTGAACAGACAGCCAGGTGTGGTGGCT 3572  
DB 1397 GAGCCACTGCACCTGCCCTATATTTTATTTAAATAAAGTTTCAGCTGGCGTGGTGGCT 1338  
QY 3573 CACGCTGTAAATCCAGCACTCTGGGAGGCTGAGGTGGGTGATCGCTTGAGCCCTCGAG 3632  
DB 1337 CATGCTGTAAATCCAGCAATTTGGGAGGCTGAGGCGGCGAGATCAC--GAGGTACAGGAG 1280  
QY 3633 TTCAAGACCAAGCTGAGCAACATGGCAAAACCCCTGTTTCT-----ATACAAAATTAG 3686  
DB 1280



Db	1279	TTGAGGCCAGCCGACCAACATGGTGAACCCCTGTCTCTACTAAAAATACAAAAATAG	1220
Qy	3687	CCGGGCATGTTGGCATGTGCTGTGGTCCAGCTACT-AGGGGCTGAGGCAGGAGAATC	3745
Db	1219	CCAGACATGTTGGTGGCTGCTGTGTAATCCAGCTACTCAGGAGGCTCAGGCAGGAGAATC	1160
Qy	3746	TTTGGAGCCAGGAGGTCAAGGTGCAAGCTGCACTGAGCAGTCTTGGCGCACTGCACTCCAGCCT	3805
Db	1159	ACTTGAACCCAGGAGGTGGAGTTGTAGTAACTGAGATTGAGCCACTGCACTCCAGCCT	1100
Qy	3806	GGGTGACAGCACAGCACTTTCCTTCAAAAAAATAAGAGAAAAAATT	3851
Db	1099	GGGCAACAGATGAGACTCCGTCTCAAAAAAAGAAAAAAGT	1054
RESULT 8			
AAS39620/c			
ID	AAS39620 standard; DNA; 32204 BP.		
XX	AAS39620;		
DT	17-DEC-2001 (first entry)		
XX	Genomic sequence #39 encoding human colon associated polypeptide.		
DE	Human; colon cancer; congenital abnormality; infection; colitis;		
KW	Inflammatory bowel disease; IBD; neoplastic disorder; gene therapy;		
KW	Intestinal inflammatory disorder; malabsorption syndrome; gastric;		
KW	Sigmoid disease; antibacterial; antiviral; antiinflammatory;		
KW	cytostatic; ds.		
XX	Homo sapiens.		
OS	WO200155302-A2.		
PN	02-AUG-2001.		
XX	17-JAN-2001; 2001WO-US01240.		
XX	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	07-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	11-JUL-2000; 2000US-0217496.		
PR	14-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225215.		
PR	14-AUG-2000; 2000US-0225266.		
PR	14-AUG-2000; 2000US-0225267.		
PR	14-AUG-2000; 2000US-0225268.		
PR	14-AUG-2000; 2000US-0225270.		
PR	14-AUG-2000; 2000US-0225447.		
PR	14-AUG-2000; 2000US-0225521.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	14-AUG-2000; 2000US-0225759.		
PR	18-AUG-2000; 2000US-0226279.		
PR	22-AUG-2000; 2000US-0226681.		
PR	22-AUG-2000; 2000US-0226686.		
PR	22-AUG-2000; 2000US-0227182.		
PR	23-AUG-2000; 2000US-0227009.		
PR	30-AUG-2000; 2000US-0228924.		
PR	01-SEP-2000; 2000US-0229287.		
PR	01-SEP-2000; 2000US-0229343.		
PR	01-SEP-2000; 2000US-0229344.		
PR	01-SEP-2000; 2000US-0229345.		
PR	05-SEP-2000; 2000US-0229509.		
PR	05-SEP-2000; 2000US-0229513.		
PR			



Db	17593	ACAGAGCAAGACTCCGCTCTTAAAAAAAATATATTTTTTAGGATATGTGATCACAGC	17534	XX	02-AUG-2001.
Qy	4773	CCAGGCATCACTTGTGCT-TTCATATACAACCACGCTGACAGCTTGTGTTGCTCTGCA	4831	XX	17-JAN-2001; 2001WO-US01324.
Db	17533	TCACTGTAACTTGAACCTCTTACTTCAAGCAGCTCCACAGCCTACAGGTGTGGCC	17474	XX	31-JAN-2000; 2000US-0179065.
Qy	4832	GCTGCAATGGGACTCTTGATTTCTTTAAGAAACTTGGGTTACCAAGAGTATTTCCACAA	4891	PR	04-FEB-2000; 2000US-0180628.
Db	17473	ACCCCTGGCTAATTTTTTAAATTTTTTGTAGAGATGGGTCATGCTATGTGCCCAGG	17414	PR	24-FEB-2000; 2000US-0184664.
Qy	4892	ATGCTATTCAAAATAGTGTATATGATATGAACACACTGTGCTAGGAGCCAGAAAAACAA	4951	PR	02-MAR-2000; 2000US-0186350.
Db	17413	CTGCTCTCAAACTCCCTGGCCTCAAGTGAGCCTCCTGCTTGGCCTCCCTAAGTGCTGGGA	17354	PR	16-MAR-2000; 2000US-0189874.
Qy	4952	GAGGAGAGAAATCAGTCAATTATGTGGGAACAATAGCAAGATATTTAGATCATTTTGA	5011	PR	17-MAR-2000; 2000US-0190076.
Db	17353	TTATAGTGGGAGC-----CACTGCATTCGGTTGCATTTCT	17318	PR	18-APR-2000; 2000US-0198123.
Qy	5012	CTAGTTAAAAAAGCAGCAGAGTACAAAATCACACATGCAATCAATTAATCCAAATCATG	5071	PR	19-MAY-2000; 2000US-0205515.
Db	17317	TTATTTAATTTACATCTTGATCTTTTCATATCAATATACATATCCTCATTTAATCTGTG	17258	PR	07-JUN-2000; 2000US-0209467.
Qy	5072	TAAATATGCCCTGTAGAAAGACTAGAGGAATAACACAAGAATCTTAAACAGTCATGTC	5131	PR	30-JUN-2000; 2000US-0215135.
Db	17257	CATAGAACACATAGTGTGGTGTAGTTTATTTAGCTATTCTTTTTTTCACGGAAA--TT	17200	PR	07-JUL-2000; 2000US-0216647.
Qy	5132	ATTAGACACTAAGCTCAATTTATTTATTTAGACACTATGATATTTGAGATTTAAAAATC	5191	PR	07-JUL-2000; 2000US-0216880.
Db	17199	ATGATTGTTTACAGTTTTTTTGCATTATACCAAGAGTGTGACAGTGAACATTTGTATAC	17140	PR	11-JUL-2000; 2000US-0217487.
Qy	5192	TTTAAATTTTAAAAATTTAGAGCTCTTCTATTTTCCATAGTATTCAGTTTGCACATGA	5251	PR	14-JUL-2000; 2000US-0218290.
Db	17139	ACCTCTTTATACATACATATGTATTTTGGAAATAGAGAAGTGGGATTTGCTGAGCTG	17080	PR	26-JUL-2000; 2000US-0220964.
Qy	5252	TCAAGTATTAATCTTTCTTT	5311	PR	14-AUG-2000; 2000US-0224518.
Db	17079	AAAGATATGATTTGTTTTTTTGGGTTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGAT	17020	PR	14-AUG-2000; 2000US-0225213.
Qy	5312	TGTTGCCCATGCTGGAGTGAATGGCATGAYCATGCTCACTGCAACCTCCACCTCCTGG	5371	PR	14-AUG-2000; 2000US-0225214.
Db	17019	TGTTGCCAGGCTGGAGTAAATGGCGTGA-CITGGCTCACTGAAACCTCTGCTCCCTA	16961	PR	14-AUG-2000; 2000US-0225267.
Qy	5372	GTTCAAGCAAGCTGTGCGCTCAGCCTCCCGGCTAGATGGATTACAGCGCCACACACC	5431	PR	14-AUG-2000; 2000US-0225270.
Db	16960	GTTCAAGCAATTTCTGCTCCTCAGCTCCCAAGTAGCTGGGATTAACAGCCACCCACC	16901	PR	14-AUG-2000; 2000US-0225757.
Qy	5432	ACACTCGGCTAA--TGTTTCTATTTTGTAGATGGGTTTCCACCATGTTGCCCAGGC	5489	PR	14-AUG-2000; 2000US-0225758.
Db	16900	ACGCCAGCTAAATTTTTTGTATTTTGTGTAGAGCGGGTTTTTACTACGTTGCCCAGGC	16841	PR	18-AUG-2000; 2000US-0226279.
Qy	5490	TGCTCTCAAACTCCTGACCTCA 5511		PR	22-AUG-2000; 2000US-0226681.
Db	16840	TGCTCTCAAACTCCTGAACCTCA 16819		PR	22-AUG-2000; 2000US-0226868.
RESULT 9					22-AUG-2000; 2000US-0227182.
AAK89019/C					23-AUG-2000; 2000US-0227009.
ID	AAK89019 standard; DNA: 32204 BP.				30-AUG-2000; 2000US-0228924.
XX	AC	AAK89019;			01-SEP-2000; 2000US-0229287.
XX	AC	AAK89019;			01-SEP-2000; 2000US-0229343.
XX	DT	05-NOV-2001 (first entry)			01-SEP-2000; 2000US-0229344.
XX	DE	Human digestive system antigen sequence SEQ ID NO: 2595.			01-SEP-2000; 2000US-0229345.
XX	DE	Human; digestive system antigen; gene therapy; cancer; appendicitis;			05-SEP-2000; 2000US-0229513.
KW	KW	ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;			06-SEP-2000; 2000US-0230437.
KW	KW	digestive system disorder; Meckel's diverticulum; ds.			08-SEP-2000; 2000US-0230438.
OS	OS	Homo sapiens.			08-SEP-2000; 2000US-0231242.
XX	XX	WO200155314-A2.			08-SEP-2000; 2000US-0231243.
PN	PN				08-SEP-2000; 2000US-0231244.



[illegible][illegible]

PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249214.
PR	01-SEP-2000;	2000US-0229345.	PR	17-NOV-2000;	2000US-0249215.
PR	05-SEP-2000;	2000US-0229509.	PR	17-NOV-2000;	2000US-0249216.
PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249218.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249244.
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000;	2000US-0232080.	PR	17-NOV-2000;	2000US-0249300.
PR	08-SEP-2000;	2000US-0232081.	PR	01-DEC-2000;	2000US-0250160.
PR	12-SEP-2000;	2000US-0231968.	PR	01-DEC-2000;	2000US-0250391.
PR	14-SEP-2000;	2000US-0232397.	PR	05-DEC-2000;	2000US-0251030.
PR	14-SEP-2000;	2000US-0232398.	PR	05-DEC-2000;	2000US-0251988.
PR	14-SEP-2000;	2000US-0232399.	PR	05-DEC-2000;	2000US-0256719.
PR	14-SEP-2000;	2000US-0232400.	PR	06-DEC-2000;	2000US-0251479.
PR	14-SEP-2000;	2000US-0232401.	PR	08-DEC-2000;	2000US-0251856.
PR	14-SEP-2000;	2000US-0233063.	PR	08-DEC-2000;	2000US-0251868.
PR	14-SEP-2000;	2000US-0233064.	PR	08-DEC-2000;	2000US-0251869.
PR	14-SEP-2000;	2000US-0233065.	PR	08-DEC-2000;	2000US-0251989.
PR	21-SEP-2000;	2000US-0234223.	PR	08-DEC-2000;	2000US-0251990.
PR	21-SEP-2000;	2000US-0234274.	PR	11-DEC-2000;	2000US-0254097.
PR	25-SEP-2000;	2000US-0234997.	PR	05-JAN-2001;	2001US-0259678.
PR	25-SEP-2000;	2000US-0234998.	XX		
PR	26-SEP-2000;	2000US-0235484.	PA	(HUMA-)	HUMAN GENOME SCI INC.
PR	27-SEP-2000;	2000US-0235834.	XX		
PR	27-SEP-2000;	2000US-0235836.	XX	Rosen CA, Barash SC, Ruben SM;	
PR	29-SEP-2000;	2000US-0236327.	PI		
PR	29-SEP-2000;	2000US-0236367.	XX	WPI: 2001-502630/55.	
PR	29-SEP-2000;	2000US-0236368.	DR		
PR	29-SEP-2000;	2000US-0236369.	XX		
PR	29-SEP-2000;	2000US-0236370.	PT	Polynucleotides encoding digestive system antigens, useful for	
PR	02-OCT-2000;	2000US-0236802.	PT	diagnosing, treating, preventing and/or prognosing disorders of the	
PR	02-OCT-2000;	2000US-0237037.	XX	digestive system, particularly cancer and cancer metastases -	
PR	02-OCT-2000;	2000US-0237038.	PS	Disclosure: SEQ ID NO 5109; 986pp; English.	
PR	02-OCT-2000;	2000US-0237039.	XX		
PR	02-OCT-2000;	2000US-0237040.	XX		
PR	13-OCT-2000;	2000US-0239935.	CC	The present invention provides the protein and coding sequences of a	
PR	13-OCT-2000;	2000US-0239937.	CC	number of human digestive system antigens. These can be used in the	
PR	20-OCT-2000;	2000US-0240960.	CC	diagnosis, treatment and prevention of digestive system disorders,	
PR	20-OCT-2000;	2000US-0241221.	CC	including cancer, Meckel's diverticulum, bacterial or parasitic	
PR	20-OCT-2000;	2000US-0241785.	CC	infections, appendicitis, Hirschsprung's disease, chronic colitis or	
PR	20-OCT-2000;	2000US-0241786.	CC	ulcerative colitis. The present sequence is a genomic DNA fragment	
PR	20-OCT-2000;	2000US-0241786.	CC	encoding a digestive system antigen of the invention.	
PR	20-OCT-2000;	2000US-0241787.	XX		
PR	20-OCT-2000;	2000US-0241808.	XX		
PR	20-OCT-2000;	2000US-0241809.	XX		
PR	20-OCT-2000;	2000US-0241826.	XX		
PR	01-NOV-2000;	2000US-0244617.	XX		
PR	08-NOV-2000;	2000US-0246474.	XX		
PR	08-NOV-2000;	2000US-0246475.	XX		
PR	08-NOV-2000;	2000US-0246476.	XX		
PR	08-NOV-2000;	2000US-0246477.	XX		
PR	08-NOV-2000;	2000US-0246478.	XX		
PR	08-NOV-2000;	2000US-0246523.	XX		
PR	08-NOV-2000;	2000US-0246524.	XX		
PR	08-NOV-2000;	2000US-0246525.	XX		
PR	08-NOV-2000;	2000US-0246526.	XX		
PR	08-NOV-2000;	2000US-0246527.	XX		
PR	08-NOV-2000;	2000US-0246528.	XX		
PR	08-NOV-2000;	2000US-0246532.	XX		
PR	08-NOV-2000;	2000US-0246609.	XX		
PR	08-NOV-2000;	2000US-0246610.	XX		
PR	08-NOV-2000;	2000US-0246611.	XX		
PR	08-NOV-2000;	2000US-0246613.	XX		
PR	17-NOV-2000;	2000US-0249207.	XX		
PR	17-NOV-2000;	2000US-0249208.	XX		
PR	17-NOV-2000;	2000US-0249209.	XX		
PR	17-NOV-2000;	2000US-0249210.	XX		
PR	17-NOV-2000;	2000US-0249211.	XX		
PR	17-NOV-2000;	2000US-0249212.	XX		
PR	17-NOV-2000;	2000US-0249213.	XX		
PR	17-NOV-2000;	2000US-0249213.	XX		

Query Match 4.1%; Score 226.8; DB 22; Length 32204;

Best Local Similarity 48.8%; Pred. No. 8.9e-29;

Matches 976; Conservative 0; Mismatches 967; Indels 59; Gaps 11;

QY	3542	AAAGTGAACAGACAGCCAGGCTGTGGTGGCTCAGCCTGTAATCCAGCACTCTGGGAGG	3601
Db	13412	AAATTAATAAATAATGCCAGTGTGGTGGCTCACACCTGTAATCCAGCACTCTGGGAGG	13471
QY	3602	CTCAGGTGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCAGGCTGAGCAACATGGCAAA	3661
Db	13472	CCAAGACCGGTGGATCACCTCAGGTGAGGATTCAGACCCAGGCTGCCAACATGGTGAA	13531
QY	3662	ACCTGTTTCTCTCTACTAAAAATACAAAAATAGCTGGGGCTGGTGGGCTGCTGTAATCC	3715
Db	13532	ACCTGTTTCTCTCTACTAAAAATACAAAAATAGCTGGGGCTGGTGGGCTGCTGTAATCC	13591
QY	3716	CAGCTACT-AGGGGGCTGAGCGAGGAGAAATCTTTGGAGCCAGGAGGTCAAGGCTGCAC	3774
Db	13592	CAGCTACTCAGGAGGCTGAGCGAGGAGAAATGGCTTGAGCCAGGAGGAGGTTCAGT	13651
QY	3775	GAGCAGTGTTCGCCCCACTGCACCTCCAGCCTGGTGACAGGACAGACACCTTGCCTCAAA	3834
Db	13652	AGCTGAGATGATGCCACTGCACCTCCAGCCTGGAAGACAGAGTGGAGACTCTGTCTCAAA	13711
QY	3835	AAATAAGAAGAAAAATTAATAATGGAACAACTACAAAGAGCTCTTCTCTCTAGATG	3894



PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0228868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-SEP-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0248524.  
PR 08-NOV-2000; 2000US-0248525.  
PR 08-NOV-2000; 2000US-0248526.  
PR 08-NOV-2000; 2000US-0248527.  
PR 08-NOV-2000; 2000US-0248528.  
PR 08-NOV-2000; 2000US-0248532.  
PR 08-NOV-2000; 2000US-0248532.  
PR 08-NOV-2000; 2000US-0248609.  
PR 08-NOV-2000; 2000US-0248610.  
PR 08-NOV-2000; 2000US-0248611.  
PR 08-NOV-2000; 2000US-0248613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 03-DEC-2000; 2000US-0251030.  
PR 03-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-457727/49.  
Isolated polypeptide for treating, preventing and/or prognosing  
disorders related to the colon and rectum including colorectal cancers  
and also for testing and detection e.g. diagnosis -  
Disclosure; SEQ ID NO: 327; 522pp + Sequence Listing; English.  
The present invention provides the protein and coding sequences of a  
number of colorectal cancer antigens. These are shown in  
AAI57547-AAI57619 and AAM38569-AAM38641. These can be used in the  
diagnosis, prevention and treatment of cancer of the colon and/or rectum.  
The present sequence is a colorectal cancer antigen genomic sequence.  
Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO





AAK89418  
ID AAK89418 standard; DNA; 28588 BP.  
XX  
AC AAK89418;  
XX  
DT 05-NOV-2001 (first entry)  
XX  
DE Human digestive system antigen genomic sequence SEQ ID NO: 2994.  
XX  
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;  
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;  
KW digestive system disorder; Meckel's diverticulum; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200155314-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001WO-US01324.  
XX  
31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.



QY	4773	CCAGGONCTACTTGTGCTTTTCATATCAACACCGCTGTACAGCTTGTGCTGTCTGCAG	4832	KW	anti-infertility.	
XX				XX		
Db	21141	GAATGGAAGACATTTTCAAAAGCGAATGGCAGTTGCCAGTTGCCAGTTCCTGGAATCTAATCCTT	21200	OS	Homo sapiens.	
XX				XX		
QY	4833	CTGCAATGGGACTCTTGATTTCTTTAAGGAACCTTGGGTTACCAAGATAT-----	4893	PN	WO200155355-A1.	
Db	21201	GGGTAGAAAGGTTATTTATCTCGGTATGTTAATTTGATTTCTAAGCAGATCTACTAGGA	21260	XX	02-AUG-2001.	
XX				XX		
QY	4884	TTCCACAAATGCTATTCAAAATAGTCTTATGATATCAAGACACTGTGCTAGGAGCCAG	4943	PF	17-JAN-2001; 2001WO-US01351.	
XX				XX		
Db	21261	TTTCAGCAGTTATCTCGAGCATCGTGTATGTATGAACATTTATTTGTGAAGTTAGTGCCTT	21320	PR	31-JAN-2000; 2000US-0179065.	
QY	4944	AAACAAAGAGGAGGAGAAATCAGTCATTA-----	4981	PR	04-FEB-2000; 2000US-0180628.	
Db	21321	CAGGTATGTTTGTGTTTCATTTATTTATTGAGCACTGCCATGTGTCAGGCTCTGTGGTGGG	21380	PR	24-FEB-2000; 2000US-0184664.	
QY	4982	CAACATAGCAAGATATTTAGATCATTTTGACTAGTTAAAGAGCAGCAGATACAAATC	5041	PR	02-MAR-2000; 2000US-0186350.	
Db	21381	TCAGGTGGAAAGAGATAGACCTGGTCTCTACTGTGTTGTTATTCCTCTAGTTTAGTGGATG	21440	PR	16-MAR-2000; 2000US-0189874.	
QY	5042	ACACATGCAATCAGTATTAATCCAAATCATGTAAATATGTCCTCTAGAAAGACTAGAGGA	5101	PR	17-MAR-2000; 2000US-0190076.	
Db	21441	CTTGGATAGTAAATCAGTTTATGCAATAGTTTCTCTGGTGGTCTGTGATTTGA	21500	PR	18-APR-2000; 2000US-0198123.	
QY	5102	ATAACACAAGATCTTAACAGTCATTTGTCATTTAGACACTAAGTCTAATATTATTATTA	5161	PR	19-MAY-2000; 2000US-0205515.	
Db	21501	AGGCTTGTGTATACCTATTCAAGATTTTAACATTAGACTTTTCAGTTGATTTCGTTATA	21560	PR	07-JUN-2000; 2000US-0209467.	
QY	5162	GACACTATGATATTGAGATTTTAAATATCTTAATATTTTAAATTTAGACGCTCTCTA	5221	PR	28-JUN-2000; 2000US-0214886.	
Db	21561	ATCTTGTCTCTGCTATTTCAGCATATAGCTGTCACTTTAATCTGATAGAGGTTTGTCCCA	21620	PR	30-JUN-2000; 2000US-0215135.	
QY	5222	TTTTTCCATAGTATTCAGGTTTGACAAATGATCAAGTATTACTCTTCTCTTTTCTTTT	5281	PR	07-JUL-2000; 2000US-0216647.	
Db	21621	GTTGAGACTTGATATTAGGTGCTAAATTCCTCAGGTTCTCTCTCTCTCTCTCTCT	21680	PR	07-JUL-2000; 2000US-0216880.	
QY	5282	TTTTTTTTTTTGAGATGGAGTTTGGTCTTGTGGCCATGCTGGAGTGAATGCATGA	5341	PR	11-JUL-2000; 2000US-0217487.	
Db	21681	TTTTTTTTTTTGAGACGGAGTTTCACTC---TTGCCAGGCTGGAGTGCATGATGAA	21737	PR	11-JUL-2000; 2000US-0217496.	
QY	5342	YCATAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCAAGCTGTGCGCTCAGCCTCC	5401	PR	14-JUL-2000; 2000US-0218290.	
Db	21738	TTTCAGCTCACTGCAACCTCTGCTCCCGGGTTCAAGTGATTTCTTCTGCTCAGCCTCC	21797	PR	26-JUL-2000; 2000US-0220963.	
QY	5402	GGGTAGATGGATTACAGGGCCACACACACACTCGGCTAATGTTTGTATTTTAGTAG	5461	PR	26-JUL-2000; 2000US-0220964.	
Db	21798	GAGTAGCTGGATTACAGTGGCCGCCACACCGCCAGCTATTTTGTATTTTAGTAA	21857	PR	14-AUG-2000; 2000US-0224518.	
QY	5462	AGATGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCTGACCTCAG	5512	PR	14-AUG-2000; 2000US-0224519.	
Db	21858	AGACGGGGTTTCACCATGTTGGCCAGTCTGGTCTCGAACTCCTGACCTCAG	21908	PR	14-AUG-2000; 2000US-0225213.	
RESULT 13					PR	14-AUG-2000; 2000US-0225266.
AAS31923					PR	14-AUG-2000; 2000US-0225267.
XX	ID AAS31923 standard; DNA: 28588 BP.				PR	14-AUG-2000; 2000US-0225270.
XX	AC				PR	14-AUG-2000; 2000US-0225447.
XX	AAS31923;				PR	14-AUG-2000; 2000US-0225757.
XX	DT				PR	14-AUG-2000; 2000US-0225758.
XX	04-DEC-2001 (first entry)				PR	14-AUG-2000; 2000US-0225759.
DE	Human liver associated genomic DNA #97.				PR	18-AUG-2000; 2000US-0226279.
KW	Liver associated protein; human; mouse; rabbit; goat; horse; cat; dog;				PR	22-AUG-2000; 2000US-0226681.
KW	chicken; sheep; immunosuppressive; antiarthritic; vasotropic;				PR	22-AUG-2000; 2000US-0226688.
KW	antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;				PR	23-AUG-2000; 2000US-0227009.
KW	cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;				PR	30-AUG-2000; 2000US-0228924.
KW	ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;				PR	01-SEP-2000; 2000US-0229287.
KW	hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;				PR	01-SEP-2000; 2000US-0229343.
KW	cerebrovascular disorder; nervous system disorder; bacterial infection;				PR	01-SEP-2000; 2000US-0229344.
KW	fungal infection; viral infection; ocular disorder; endocrine disorder;				PR	05-SEP-2000; 2000US-0229509.
KW	gastrintestinal disorder; renal disorder; respiratory disorder;				PR	05-SEP-2000; 2000US-0229513.
KW	wound healing; skin aging; organ transplantation; tissue regeneration;				PR	06-SEP-2000; 2000US-0230437.
					PR	06-SEP-2000; 2000US-0230438.
					PR	08-SEP-2000; 2000US-0231242.
					PR	08-SEP-2000; 2000US-0231243.
					PR	08-SEP-2000; 2000US-0231244.
					PR	08-SEP-2000; 2000US-0231413.
					PR	08-SEP-2000; 2000US-0231414.
					PR	08-SEP-2000; 2000US-0232080.
					PR	08-SEP-2000; 2000US-0232081.
					PR	12-SEP-2000; 2000US-0231968.
					PR	14-SEP-2000; 2000US-0232397.
					PR	14-SEP-2000; 2000US-0232398.
					PR	14-SEP-2000; 2000US-0232399.
					PR	14-SEP-2000; 2000US-0232400.
					PR	14-SEP-2000; 2000US-0232401.
					PR	14-SEP-2000; 2000US-0233063.
					PR	14-SEP-2000; 2000US-0233064.
					PR	14-SEP-2000; 2000US-0233065.
					PR	21-SEP-2000; 2000US-0234223.
					PR	21-SEP-2000; 2000US-0234274.
					PR	25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 12-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.

XX (HUMA-) HUMAN GENOME SCI INC.  
PA Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-457728/49.  
XX Isolated nucleic acid molecule encoding a human liver related protein  
is used in preventing, treating or ameliorating disorders of the liver  
particularly cancer of the liver -  
XX Claim 1; SEQ ID No 399; 526pp; English.  
XX Sequences AAS31827-AAS32182 represent genomic DNA molecules, which encode  
the liver associated polypeptides of the invention. Liver associated  
polypeptides and their associated polynucleotides are useful in the  
diagnosis, treatment and prevention of various types of disorders in e.g.  
humans, mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A  
pathological condition can be determined by detecting the presence or  
absence of a mutation in a liver associated polynucleotide. The treatable  
disorders include autoimmune diseases such as rheumatoid arthritis,  
hyperproliferative disorders such as neoplasms of the breast or liver,  
cardiovascular disorders such as cardiac arrest, cerebrovascular  
disorders such as cerebral ischaemia, nervous system disorders such as  
Alzheimer's disease, infections caused by bacteria, viruses and fungi,  
ocular disorders such as corneal infection, endocrine disorders such as  
premature labour and infertility, gastrointestinal disorders such as  
Crohn's disease, renal disorders such as glomerulonephritis and  
respiratory disorders such as asthma and pleurisy. The polypeptides can  
also be used to aid wound healing, to prevent skin aging due to sunburn,  
to maintain organs before transplantation, to regenerate tissues and in  
chemotaxis.  
XX Note: The sequence data for this patent did not form part of the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

Query Match 4.1%; Score 225.4; DB 22; Length 28588;  
Best Local Similarity 48.3%; Pred No. 1.5e-28;  
Matches 1154; Conservative 1; Mismatches 1142; Indels 94; Gaps 15;  
QY 3213 TTTATGTTTTTTTATGTTTTTTCATTTTAAAGACAATCTCACCTGTACCCAGGCTGG 3272  
DB 19521 TTTTATTTTATTTATTAATTTTGTGAGATGGAGTTTCTGTTGCCAGGCGGG 19580  
QY 3273 AGTCAGTGGTGCATATCATAGCTTTCGACGTCCTGAACTCCTGGGCTCAAGCATCTCTC 3332  
DB 19581 AGTGAAGTGGCGTGATCTCAGCTCACTGCAACCTCCACCTCCTGGGTTCAAGCGATTCTC 19640  
QY 3333 CTGCCTTGGCCTCCCAAGTGTGGGAT---ACAGTCATGAGCCACTGCATCTGGCCTAG 3389  
DB 19641 CTGCCTCAGCCTCTGAGTAGCTGGGATTACAGGTGCCACCACCATGCCAGCTAATTT 19700  
QY 3390 GATCCATTTAGATTAATAATATGCAATTTTAAATTTTAAATTAATATG---CTAATTTT 3445  
DB 19701 TTTATTTTGTAGACAGCGGTTTCCACGCTGTGACAGGCTGGTCTCAAACTCCTG 19760  
QY 3446 ACCTTATGTAAT-----GTGTATCTGGTAAATAAATCTAGTTTGTCCCTAAAGTTTA 3498  
DB 19761 ACCTCAGGTGATCCACCTTGGCCTCCCAAGTGTGGGATTACAGGCTAGAGCCAC 19820  
QY 3499 AAGTGTTCCTCAATAAGCTTC-----ATGTACGTGAGGGGAGACATTTAAAGTGAACA 3552  
DB 19821 CGTCCCTGGCCAAAAAAATTTTCTGATTTTAGAGGAAGTGTTATATAAATAGTATT 19880  
QY 3553 GACAGCCAGGTGTGGTGGCTCAGCGCTGTAATCCAGCACTCTGGAGGCTGAGGTGGGT 3612  
DB 19881 GCAGGCTGGGCACGGGTGGCTCATGCTCTGTAATCCAGCACATTTGGAGGCTGAGGCGAGT 19940  
QY 3613 GGATCGCTTGAGCCCTGGAGTTCAAGACCGCTGAGCAACATGCGCAAAACCCCTGTTTCT 3672  
DB 19941 GGATCAGTTGAGCTTAGGAGTTTGAGATCGGCTCGAGCAACATGCGCAAAACCCCTGTTTCT 20000

QY	3673	ATAACAAA-----AATTAGCGGCATGGTGGCATGTGCTGTGGTCCAGCTACTA	3724	Db	21081	CAGTTCCACTCTCTATTGATTGTTACCTCTGCAATCGAGGCTCGAAGACTCTACATGTCC	21140
Db	20001	ACCAAAATACAAACCAATTAGCCAGGTGCGATGCGATGTGCTGTGGTCCAGATACTC	20060	QY	4773	CCAGGCATCAGTTGCTTTTCATATCAACACGCTGTACACGCTGTGTTGCTGTCTGCAG	4832
QY	3725	GGGGCTGAGGAGAGAAATCTTTGGAGCCAGGAGGTCAAGGTCGACGTGACAGTGTCT	3784	Db	21141	GAATGAAAGAGATTTCAAAACGGAAGGAGTGGCCAGTGTCTGGAAATCTAATCCTT	21200
Db	20061	AGAGGCTGAGGAGGAGATTACCTGAGCCTGGAGGCGAAGGTTGCAGTGAGCTGACAT	20120	QY	4833	CTGCAATGGGGACTCTTGATTTCTTTAAGGAAACTGGGTTACCAGAGPAT-----	4883
QY	3785	TGCGCCACTGCACTCCAGCCTGGGTGACAGGACCAGACCTTGCCTCAAAAAATAGAG	3844	Db	21201	GGGTAGAAAAGGTTATTATCTGCTGATGTTAATTTGATTCTTAAGACAGATCTACTAGGA	21260
Db	20121	CCTGCACTGCAATCCAGCCTGGGCAACAGAGTGAATGCTCTCAAAAAACAAAAAC	20180	QY	4884	TTCCCAAAATGCTATTCAAATTTAGTCCTTATCATATGCAAGACATGCTCTAGGAGCCAG	4943
QY	3845	AAAAATTAATAATAATGAAACAACTACAAAGAGC-----TGTTGTCC	3888	Db	21261	TTTCAGCAGTTATCTCTGAGCATCGTGTATTGTAACATTTATTGTAAGTTAGGTGCTTT	21320
Db	20181	AAAACGGTTATAGCATAGCAGGCTTTATGATATGATATAAGAAATGAGTTTATTAT	20240	QY	4944	AAAAACAGAGGAGAGAAATCAGTCATTA-----TGTTGGGAA	4981
QY	3889	TAGATGAGTACTTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGCTGT	3948	Db	21321	CACGTATTGTTTTGTTTCATTTATTATTAGCAGCTGCCATGTGTACGGCTCTGTGTGG	21380
Db	20241	TTTCATGCTCCATCTTAATATAATCATAAATTTGATAAGGCCCTAGTGAATTTTGTTC	20300	QY	4982	CAACATAGCAAGATATTTAGATCATTTTGACTAGTTTAAAAAGACGACGAGTACAAAATC	5041
QY	3949	CACCTGCACATATTAATAAATATCAATCTCAATGTATATCCACACAAAGACTGGTAC	4008	Db	21381	TCAGGTGGAAAGAGATAGACCTGGTCTCTACTGTTGTTATGCTCTAGTTAGTGGATG	21440
Db	20301	AGACTTAATGGTTTTTTTTCTTTTAAAAATTACATTTGAAATGAGGCTTTTCATATAC	20360	QY	5042	ACACATGCAATCAGTATATATCCAAATCATGTAATATGTCGCTGTAGAAAGACTAGAGGA	5101
QY	4009	GTGAATGTTTCATAGTAC-CTTTATTACAAAAACCCCAAGTAGAGACTATCCAAATATCC	4067	Db	21441	CTTGATAGTAGAAAATCAGTTTATGTCATTTAGTTGTTCTGCTGGTGGTCTGTGATTGA	21500
Db	20361	TATACCTTTTATATCACGAATGATGTTAAATTAATAGCACTTACAGTATAAAGTTAC	20420	QY	5102	ATAAACACAAAGATCTTAACAGTCATTGTCATTAGACACTAAGTCTAAATTTATTATTAA	5161
QY	4068	ATCAACAAGTGAACAAATTAAGCAAAATGTGCTATATCCATGCAATGGAATACACCCCTGC	4127	Db	21501	AGCTTGTGTATACCTATTCAAAGATTTTAAACATTAGACTTTTTCAGTTGATTGCTTATA	21560
Db	20421	ATTTTCAATAGATGTCTAAACACCATTTCCTATTATTTACATTTGCTTAGTTTATCTCC	20480	QY	5162	GACACTATCATATTTGAGATTTAAAAAATCTTTTAAATATTTTAAAAATTTAGAGCTCTCTA	5221
QY	4128	AGTACAAAGGAGAGC---TACTTGGGGATGAATCCCAAGTCATGACGCTGAATGAA	4183	Db	21561	ATCTTCTCTCTGCTGCTATTTCAGCATTTAGCTGCACCTTAACTCTGATAGAGGTTTGTCC	21620
Db	20481	TAGTTAAAGATTAGCCATTGTAACCTTGACAGGTAAATATTAACATGATTAGGATATA	20540	QY	5222	TTTTTCCATAGATTCAAGTTTGACAAATGATCAAGTATTACTCTTTCTTTTCTTTTCTTT	5281
QY	4184	AGAGTCAGACTGAAGGAGAGATAATGTATGCCATACGAAAT-TCAGAAAATGAAAT	4242	Db	21621	GTTGAGACTTGATATTAGGTGCTAAATTTCCCTGAGTTCCTCTCTCTCTCTCTCTCTCT	21680
Db	20541	AATGTCTCCACAAGGAAACTGAAACTTATTGCTTGATAAGTGTCTTTGGTTTATAAA	20600	QY	5282	TTTTTTTTTTTGGAGATGGATTTTGGTCTTGTGTGCCCATGCTGGAGTGGATGGCATGA	5341
QY	4243	AACCTATAGTACAGAAAGCAATCAGGCGAGCAGCATAGAGCTCACACCTGTAATCCCAG	4302	Db	21681	TTTTTTTTTTTGGAGCGGAGTTTTCACCTC---TTGCCAGGCTGGAGTGCATGATGCA	21737
Db	20601	AGTAGAAATGAAGTATATCTCGAGGCATGTGAATCATTTGACCGCTTTGCACAAAG	20660	QY	5342	YCATAGCTCACTGCACACCTCCACCTCCTGGGTTCAAGCAAGCTGTGCGCTCAGCCTCCC	5401
QY	4303	CACTTTGAAGGCGACGTGGGAAGATTGCTAGAACCTCAGGAGTTCAAGACAGCCTGGGC	4362	Db	21738	TTTTCAGCTCACTGCAACCTCTGCCCTCCGGGTTCAAGTGAATCTCTTGGCTCAGCCTCCC	21797
Db	20661	GCTAAATAGNATATTAGTAGAATATTTCAGTAATATAGGACTTGAACCTGCTGTTTCTC	20720	QY	5402	GGGTAGATGGGATTACAGGCGCCACACACACTCGGCTAAATGTTGTATTTTATTAGTAG	5461
QY	4363	AACACAGTGAATCCATTCTCCACAAAATGGGAAAAAAGAAAGCAATCAGTGGTTG	4422	Db	21798	GAGTAGTGGGATTACAGGTGCGCCGCCACCCAGCTATTTTGTATTTTATTAGTAA	21857
Db	20721	TGTATGTTAAATAAATTTTCAAGAAATACAAGTTTAGCTATGAACRAATTTATGATTT	20780	QY	5462	AGATGGGGTTTCCACCATGTTGGCCAGGCTGCTCAAACTCCCTGACCTCAG	5512
QY	4423	TCCTGTGGGAGGAGGACTCCAAGAGGGAGA--AGCTCTGTGGGTGAGGGTG	4479	Db	21858	AGAGGGGTTTCCACCATGTTGGCCAGTCTGGTCTCGAACTCCTGACCTCAG	21908
Db	20781	TATTAATAATCATAAGAGGACTACTTATCTTCATATTGAGAAGTTTATATCATTTGTGCT	20840	RESULT 14			
QY	4480	GTCAFTCAGGTTCTGTATCTGACTGTGTGAGCAGTTTGGGTGTTTACATCCAAAATA	4539	ABN90278			
Db	20841	TTTACTTAGTTTTTCAGCATATTTGCTTCCCTCATTTTAAAGTAAATTTCTCTAAGCATA	20900	ID	ABN90278	standard; DNA; 28588 BP.	
QY	4540	TTGCTAGAAATATGCAATCTTAATATGGGTGAGTGTACTGTATGTAATTAATTAATG	4599	XX	ABN90278;		
Db	20901	TTCTATGCTTTGTTCTTCTCTGATTTGATATATATTTCCTCCTAGAACTCAAAC	20960	AC	24-JUL-2002	(first entry)	
QY	4600	TAAGAAAAATAATGTGAAGAAAGTTTCAATTTCTTCCAGCAAACTGTTTCAAAAT	4659	XX	Human liver antigen HLDRI94	genomic sequence, SEQ ID NO:399.	
Db	20961	ATTGAGCTTTACACCTAGCTTCTGAGAGTCACTGTAATGAAAGAGAACCCAGATTAAT	21020	XX	Human; liver antigen; liver disorder; hepatic disorder; infection;		
QY	4660	TCCTGAGCCCTTTTACTTCGAAATTTCTCTCACT-TCGTCCCGGTACCATTAGTGACAG	4718	KW	hepatitis; viral; parasitic; bacterial; fungal; inflammatory condition;		
Db	21021	TCCTGGCCATACCCCTGCTTCAAACTTATCCAGGATGTGTTCAATTTTTCAGCTCTTGAG	21080	KW	cirrhosis; granulomatous hepatitis; toxin damage; drug damage;		
QY	4719	CACTAGCTCCACAAATGGGATAAATGCATTTCTTGAAAAAGACTAGGGACAAAA-----T	4772	KW	autoimmune disease; Wilson's disease; primary biliary cirrhosis;		
				KW	neoplastic disorder; cancer; tumour; portal hypertension;		
				KW	gastrointestinal disorder; hepatitis; drug screening; gene therapy;		
				KW	chromosome mapping; forensic analysis; antibody preparation;		





Db 20001 ACCAAAAATACAAACCAATAGCCAGGTGAGTGGCATGTGCTGTGGTCCCGAGATACTC.20060  
 Qy 3725 GGGGGCTAGGCGAGGAGAACTTTTGGAGCCGAGAGTCAAGGCTGCACTGAGCAGTGCT 3784  
 Db 20061 AGAGGCTGAGCGAGGAGGATTAACCTGAGCCTGGGAGCGAAGGTGTGCACTGAGTGACAT 20120  
 Qy 3785 TGGGCCACTGCATCCAGCCTGGGTGACAGGACGAGACCTTGCTCAAAAAATAAGAAG 3844  
 Db 20121 CGTGCCACTGCAATCCAGCCTGGGCAACAGAGTGAGATCCTGTCTCAAAAAACAAAAAC 20180  
 Qy 3845 AAAAAATTAATAATTAATTAAGAAACAACTACAAAGAGC-----TGTGTGTC 3888  
 Db 20181 AAAAAAGGTATAGCATAGCAAGGCTTATGATATGATATAGAAATGAGTTTTTTTATT 20240  
 Qy 3889 TAGATGAGCTACTTAGTTAGGCTGATATTTTGGTATTTAACTTTTAAAGTCAGGGTCTGT 3948  
 Db 20241 TTTATGTCCTCACTTAATTAATTAATCAATTTTGAATAAGCCCTAGTGAATTTTAGTTTCA 20300  
 Qy 3949 CAGCTGCACCTACATTAATAATCAATCTCAATGTATATCCACACAAAGACTGGTAC 4008  
 Db 20301 AGACTTAATGGTTTTTTTTTTTAAATTTACATTTGAATGAGGCGTTTCATATAC 20360  
 Qy 4009 GTGAATGTTTCATAGTAC-CTTTTATTCACAAACCCCAAGTAGAGACTATCCAAATATCC 4067  
 Db 20361 TATACCTTTTATATCAGGAATGCTTAATTTTAATAGCCTTACAGTATAACTTAC 20420  
 Qy 4068 ATCAACAAGTGAACAATTAACAATAATGTGTATATCCATGCAATGGAATACCAACCCCTGC 4127  
 Db 20421 ATTTTCAATAGATGTCTAATAACACCAATTTCCCTATTTTACATTTGCTTAGTTTATCTCCAGTA 20480  
 Qy 4128 AGTACAAAGGAAGAAGC----TACTTGGGGATGAATCCCAAGCTATGACGCTAAATGAA 4183  
 Db 20481 TAGTTAAAGATTAAAGCCATTGTAACCTGACAGTTAAATATAAACATGATTAGGAATAAA 20540  
 Qy 4184 AGAGTCAGACATGAAGGAGAGATATGATGCCATACGAAAT-TCTAGAAATGAAAGT 4242  
 Db 20541 AATGTCTCCACAGGGAACCTGAACTTATTGCTTGATAAGTGCTTGGTTTATATAA 20600  
 Qy 4243 AACTTATAGTTACAGAAACCAATCAGGCGAGCATAGAGGCTCAGCCTGTAATCCGAG 4302  
 Db 20601 AGTAGAATGAGTATATATCCTGGAGGCATGTGAATATCATTTGACCGCTTTGCAACAAG 20660  
 Qy 4303 CACTTTGAGAGCCACGCTGGGAAGATTGCTAGAACTCAGAGGTTCAAGACCAAGCCTGGCG 4362  
 Db 20661 GCTAAATAGAAATATTAGTAGAATATTTCAGTAAATATAGGACTTGAAGTCTGCTTTTCTC 20720  
 Qy 4363 RACACAGTCAAACTCCATCTCCACAAATGGGAAAAAAGAAAGCAATCAGTGGTGG 4422  
 Db 20721 TGTATGTTAAAAATAATTTTCAAGAGAAATACAAGGTTAGGCTATGAACAATTTATGATT 20780  
 Qy 4423 TCCTGTGGGGAGGGAAGGACTGCAAGAGGGAAGA---AGCTCTGGTGGGTGAGGGTG 4479  
 Db 20781 TATTAAATCATAGAAAGGACTACTTCTTCATATTTGGAAGGTTTATATCATTTGTGCT 20840  
 Qy 4480 GTGATTCAGGTTCTGTATCTCTGACTGTGTGAGCAGTTTGGGTGTTTACATCCAAAAATA 4539  
 Db 20841 TTTACTTAGTTTTCAGCATATTTGCTTCCCTCATTTTAAAGTAAATTTCTTAGCATATA 20900  
 Qy 4540 TTCGTAGAATATGATCTTAATGGGGTGGAGTTTACTGTATGTAAATATACCTCAATG 4599  
 Db 20901 TTTCTATGCTTTGTTCTTCTGCTGTGATTGATATATTTTCTCTCTAGAACTCAAA 20960  
 Qy 4600 TAAGAAAAAATAATGTGTAGAAAGATTTCATTTCTTTGCGCAGCAAAAGCTTATTCAAAT 4659  
 Db 20961 ATTGAGCTTTACACTAGCTTCTGAGAGTCACTGTAATGGAAAGAGAGCCAGCATAAAT 21020  
 Qy 4660 TCCTGAGCCCTTTACTTCGCAAAATCTCTGCACT-TCTGCCCCGTACCAATTAGGTGACAG 4718  
 Db 21021 TCTTGGCCATACCTCGCTTCAAACTTATCCAGGATGTGTTCATTTTGCAGCTCTTGGAG 21080  
 Qy 4719 CACTAGCTCCCAAAATTTGATTAATGCAATTTCTGGAAGAGACTAGGACAAAA-----T 4772  
 Db 21081 CAGTTCCATCTCCTATTGATTGTTTACCTCTGCAATTCGAGGCTCGAAGACTCTACATGTC 21140

Qy 4773 CCAGGCATCACTTGTGCTTTTCATATCAACCACGCTGTACAGCTTGTGCTGCTGTCAG 4832  
 Db 21141 GAATGGAAGAGCAATTTCAAAAACGGAATGGCAGTTGCCAGTTCCCTGGAATCTAACTTT 21200  
 Qy 4833 CTGCAATGGGACTCTTGTATTTCTTTAAAGAAACTTTGGGTTACACAGATAT----- 4883  
 Db 21201 GGGTAGAAAGGTTTATTTATCTCTGGTATGTTAATTTGATTTCTTAAGCAGATCTACTAGGA 21260  
 Qy 4884 TTCCACAATGCTATTTCAAAATAGTCTTATGATATGCAAGACACTGTCTAGGAGCCAG 4943  
 Db 21261 TTTCAGCAGTATCTCTGAGCATCGTGTATGTAATGTAACATTTATTTGTAAGTGTAGTCTTT 21320  
 Qy 4944 AAAAAAGAGGAGGAGAAATCAGTCATTA-----TGTCGGAA 4981  
 Db 21321 CAGGTATGTTTTTGTTCATTTTATTTAGCAGCTGCCATGTGTACAGGCTCTGTGTGGG 21380  
 Qy 4982 CAACATAGCAAGATATTTAGATCATTTTGCATAGTTTAAAAAGACAGCAGACAGTACAAAATC 5041  
 Db 21381 TCAGGTGGAAGAGATAGACCTGGTCTCTACTGTGTGATTTGCCCTACTTTAGTGGATG 21440  
 Qy 5042 ACACATGCAATCAGTATATATCCAAATCATGTAATATGTCCTGTAGAAAGACTAGAGGA 5101  
 Db 21441 CTTGGATAGTAGAAAAATCAGTTTATGCAATTTAGTTGTTCTGCTGGTGGTCTGTGATTGA 21500  
 Qy 5102 ATAAACACAAAGAAATCTTAACAGTCTATTCATTTAGACACACTAAGTCTAATTTATTTATTA 5161  
 Db 21501 AGGCTTGTGTATACCTATTTCAAGATTTTAAAGATTTTACATTTAGACTTTTTCAGTTGATTGCTTATA 21560  
 Qy 5162 GACATATGATATTTGAGATTTTAAAAATCTTTAATATTTTAAAAATTTAGAGCTCTTCTTA 5221  
 Db 21561 ATCTTGTCTGTCTGCTATTTTACAGATTTAGCTGTCACTTTTAACTGTATAGAGGTTTGTCCCA 21620  
 Qy 5222 TTTTTCCTAGTATTTCAAGTTTGGACAATGATCAAGTATTAAGTCTTTCTTTTCTTTTCTTTT 5281  
 Db 21621 GTTGAGACTTGATATTTAGGTGCTAAATTTCCCTGAGGTTCTCTCTCTCTCTCTCTCTCTCT 21680  
 Qy 5282 TTTTTTTTTTTGAGATGGAGTTTGGTCTTTTGGCCCATGCTGGAGTGAATGCATGA 5341  
 Db 21681 TTTTTTTTTTTTGACAGCGAGTTTTCACATC---TTGCCCGAGGCTGGAGTGAATGATGCA 21737  
 Qy 5342 YCATAGCTCACTGCAACCTCCACCTCTCGGGTTCAAGCAAAAGCTGTCCGCTCAGGCTCCC 5401  
 Db 21738 TTTTCCAGCTCACTGCAACCTCTGCGCTCCCGGTTCAAGTGAATGATGATGCA 21797  
 Qy 5402 GGGTAGATGGGATTACAGCGCCGCCACCACTGCGGCTAATGTTGCTATTTTCTAGTAG 5461  
 Db 21798 GAGTAGCTGGGATTACAGGTCGCCGCCACCAAGCCAGCTATTTTTTGTATTTTAGTAA 21857  
 Qy 5462 AGATGGGGTTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGACCTCAG 5512  
 Db 21858 AGACGGGGTTTTCACCATGTTGGCCAGTCTGTGCTCGAACTCCTGACCTCAG 21908

RESULT 15

AAAL05277 standard; DNA; 10855 BP.

AC AAAL05277;

DT 21-NOV-2001 (first entry)

DE Human reproductive system related antigen DNA SEQ ID NO: 7965.

XX Human; reproductive system related antigen; reproductive system disorder;  
 KW cancer; gene therapy; ds.

OS Homo sapiens.

PN WO200155320-A2.

PD 02-AUG-2001.

XX



```
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
(HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
DR
XX Isolated nucleic acid molecule encoding a reproductive system antigen
PT
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 18, 2003, 23:34:44 ; Search time 7435.16 Seconds  
(without alignments)  
12010.756 Million cell updates/sec

Title: US-09-966-880A-9  
Perfect score: 5514  
Sequence: 1 acacagcaatacatgtgcc.....tcaaacctcgtacctcagag 5514

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_htc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_htc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	291.4	5.3	293	17 A0042682	A0042682 CIT-HSP-2
2	199.4	3.6	694	17 AG048607	AG048607 Pan trogl
3	199.2	3.6	373	17 A0045714	A0045714 RPII11-35
4	199.2	3.6	408	17 A0046391	A0046391 RPII11-35
5	196.4	3.6	403	13 BG941633	BG941633 ax15d09.x
6	196	3.6	689	14 BM990649	BM990649 UI-H-DIO-

C	7	195.2	3.5	894	14 BQ932123	BQ932123 AGENCOURT
C	8	195.2	3.5	1769	11 AF130079	AF130079 Homo sapi
C	9	194.8	3.5	1027	12 BG539940	BG539940 602567448
C	10	194.4	3.5	1016	14 BQ674146	BQ674146 AGENCOURT
C	11	194.2	3.5	528	17 A0734866	A0734866 HS_3051_A
C	12	194.2	3.5	532	17 A0356384	A0356384 CITBI-EI-
C	13	194	3.5	451	14 BM989249	BM989249 UI-H-DPO-
C	14	194	3.5	885	12 BF575954	BF575954 602132930
C	15	193.4	3.5	778	17 A0743399	A0743399 HS_5387_B
C	16	193.2	3.5	801	9 AUI18632	AUI18632 AUI18632
C	17	192.8	3.5	550	9 AIT754064	AIT754064 cr19c04.x
C	18	192.8	3.5	916	14 BQ673918	BQ673918 AGENCOURT
C	19	192.2	3.5	529	12 BF897850	BF897850 QV1-MT022
C	20	192	3.5	370	12 BF928499	BF928499 IL2-NT020
C	21	192	3.5	513	14 BM768638	BM768638 K-EST0051
C	22	191.8	3.5	424	17 A0807427	A0807427 HS_3177_A
C	23	191.8	3.5	468	17 A0530817	A0530817 RPII-11-3
C	24	191.8	3.5	654	14 BM999209	BM999209 UI-H-DIO-
C	25	191.8	3.5	764	12 BG428850	BG428850 602494635
C	26	191.6	3.5	554	9 AL713338	AL713338 DKF2p686p
C	27	191.6	3.5	872	14 BQ674797	BQ674797 AGENCOURT
C	28	191.2	3.5	810	14 BQ670979	BQ670979 AGENCOURT
C	29	191	3.5	440	10 BE677227	BE677227 7d82b10.x
C	30	191	3.5	504	10 AV755268	AV755268 AV755268
C	31	190.8	3.5	402	17 A0087306	A0087306 HS_2191_B
C	32	190.6	3.5	279	14 C15966	C15966 C15966 Clon
C	33	190.6	3.5	425	17 A0513367	A0513367 HS_5073_B
C	34	190.6	3.5	796	17 A0490320	A0490320 RPII-11-2
C	35	190.4	3.5	654	14 BQ188600	BQ188600 UI-E-EJ1-
C	36	190	3.4	670	17 AG104085	AG104085 Pan trogl
C	37	189.6	3.4	394	10 BE077105	BE077105 RC5-BT060
C	38	189.6	3.4	667	10 AV710119	AV710119 AV710119
C	39	189.6	3.4	672	17 AG102362	AG102362 Pan trogl
C	40	189.4	3.4	380	12 BF803038	BF803038 CMI-C1009
C	41	189.4	3.4	443	9 AA595661	AA595661 n149h02.s
C	42	189.4	3.4	599	9 AL702649	AL702649 DKF2p6860
C	43	189.2	3.4	336	9 AI310670	AI310670 ta46g10.x
C	44	189.2	3.4	826	14 BQ438665	BQ438665 AGENCOURT
C	45	189	3.4	329	9 AI358712	AI358712 qx14c02.x

#### ALIGNMENTS

RESULT 1  
A0042682  
LOCUS CIT-HSP-2326M11.TR CIT-HSP Homo sapiens genomic clone 2326M11, DNA  
DEFINITION A0042682 293 bp linear GSS 14-JUL-1998  
sequence.  
ACCESSION A0042682  
VERSION A0042682.1 GI:3310067  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 293)  
AUTHORS Adams,K., Berry,K., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,  
Golden,K., Simon,M., and Venter,J.C.  
TITLE Use of a random BAC End Sequence Database for Sequence-Ready Map  
Building (1998)  
JOURNAL Unpublished (1998)  
COMMENT Other\_GSSs: CIT-HSP-2326M11.TV  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:



## FEATURES

BASE COUNT	60 a
ORIGIN	
Query Match	Score 199.2; DB 17;
Local Similarity	Pred. No. 6.7e-13;
Query Match	Mismatches 34;
Local Similarity	Indels 0; Gaps 0;
Query Match	Length 373;
Local Similarity	

[illegible]

```

RESULT 4
AQQ046391
AQQ046391
LOCUS
DEFINITION
ACCESSION
AQQ046391
VERSION
AQQ046391.1
KEYWORDS
GI:315318
SOURCE
Human.
Organism
Homo sapiens
Chordata; Craniata; Vertebrata; Euteleostomi;
Eukaryota; Metazoa; Primates; Catarrhini; Hominidae; Homo.
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 408)
Adams, W. D., Rounsley, S. D., Zhao, S., Field, C. E., Bass, S., Linher, K.,
Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., de Jong, P. and
Venter, J. C.
Venter, J. C.
Other_GSSs:
Unpublished (1998)
Other_GSSs:
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
7912 Medical Center Dr., Rockville, MD 20850, USA
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 0208
Email: mradams@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
(http://bacpac.med.buffalo.edu/ordering) or
BAC PAC Resources (http://bacpac.med.buffalo.edu/ordering). BAC end search page:
research Genetics (info@resgen.com). BAC end search page:

```

FEATURES source

```

/clone_lib="RPC1-11
/sex="Male"
/sex="Male" "Lymphocytes"
/cell_type="PBACE3.6; Site_1: ECORI; Site_2: ECORI;
/note="vector: PBACE3.6; Site_1: ECORI; Site_2: ECORI;
/note="vector: PBACE3.6; Site_1: ECORI; Site_2: ECORI;
RPC111 Human Male BAC Library 149 t
90 g
RPC102 C
67 a

```

[illegible]

RESULT 5	403 bp mRNA
BG941633	lincRNA
DEFINITION	403 bp mRNA
ACCESSION	403 bp mRNA
VERSION	403 bp mRNA
KEYWORDS	403 bp mRNA
SOURCE	403 bp mRNA
ORGANISM	403 bp mRNA
REFERENCE	403 bp mRNA
AUTHORS	403 bp mRNA

1 (bases 1-1000) of the human erythroid cell line Gubli.A.N., Njoroge.J.M., *Blood* 84:1000-1005 (1995).  
**REFERENCE**  
**AUTHORS** Gubli.A.N., Njoroge.J.M., *Blood* 84:1000-1005 (1995).  
**TITLE** Gene expression in proliferating erythroid cells.  
**JOURNAL** *Genomics* 59 (2), 168-177 (1999).  
**MEDLINE** 99393981  
**COMMENT** Contact: Jeffery L. Miller  
 Laboratory of Chemical Biology  
 National Institute of Diabetes and Digestive and Kidney Diseases  
 Building 10, Room 9B17, National Institutes of Health, Bethesda,  
 Building 10, Room 9B17, National Institutes of Health  
 20892, USA  
 Tel: 301 402 2373  
 Tel: 301 435 5148  
 Fax: 301 435 5148  
 Email: jlm1@nih.gov  
 DNA Sequencing and analyses by National Institutes of Health  
 Sequencing Center (NISC).  
 Intramural Sequencing Center (NISC).  
 plate: 15 row: d column: 09  
 primer: 21M13 forward primer (ABI).  
 primer: 21M13 reverse primer (ABI).

```

FEATURES
source
Location/
1. .403
/organism="Homo sapiens"

```



HVGF INCFN			
BASE COUNT	431 a	288 c	703 t
ORIGIN			

	Query Match	Score	DB 11;	Length	DB 1769;
	Best Local Similarity	80.5%;	Pred. No.	5.9e-13;	
	Matches	227; Conservative	1; Mismatches	54;	Indels 0; Gaps 0;
QY	5231	ACTATTCAAGCTTTGACAATGATCAAGATATTAACCTCTTCCTTTTTTTTTTTTTTTTTTTTTT		5290	
Db	594	AATTTTTCAGCTCTTTCATGGTTGTGTGTTGTTGTTTATTTTATTTTATTTTATTT		653	
QY	5291	TTTGAGATGGAGTTTGGTCTTGTTGCCATGCTGGAGTGGAAATGGCATGAACAGCTC		5350	
Db	654	TTTGAGATGGAGTTTCCGCTCTTGTTGCCATGGCGGAGTCAATGGCGTGATCTGGCT		713	
QY	5351	ACTGC AACCTCCACTCCTCGGGTTCAAAGCAAAGCTGTCGCCTCAGCCTCCCGGGTAGATG		5410	
Db	714	ACTGCAGCCTCCGCGCTCCTGGGTTCACAGCAATTCCTCTGTCTCAGCCTCCCGAGTAGCTG		773	
QY	5411	GGATTACAGGCGCCACACACCACACTCGGCTAATGTTTGTATTTTAGTAGACATGGGT		5470	
Db	774	GGATTACAGTGGCTGCCACACAGCCAGCTAATTTCTGTATTTTGTAGTAGACAGGGT		833	
QY	5471	TTCAACATGTGGCAGGCTGGTCTCAAACTCCAGCTCAG		5512	
Db	834	TTCATCATATGGTGCAGGCTGGTCTTGAACTCCTGACCTCAG		875	

RESULT 9	
BG539940/c	
LOCUS	1027 bp mRNA linear EST 03-APR-2001
DEFINITION	G02567448F1 NIH_MGC_77 Homo sapiens cDNA clone IMAGE:4692322 5',
	mRNA sequence.
ACCESSION	BG539940
VERSION	BG539940.1 GI:13532173
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgabbs-f@mail.nih.gov">cgabbs-f@mail.nih.gov</a> Tissue procurement: CLONTECH Laboratories, Inc. cDNA Library preparation: CLONTECH Laboratories, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: L1CMI1512 row: 0 column: 11 High quality sequence stop: 322.

```

FEATURES
source
  high quality sequence stop: 322.
  location/Qualifiers
    1. 1027
      organism="Homo sapiens"
      db_xref="taxon:9606"
      /clone="IMAGE:4692322"
      /clone_lib="NIH_MGC.77"
      /lab_host="DH10B (T1 phase-resistant)"
      /note="organ: lung; Vector: pDNR-LIB (Clontech); Site:1:
      SfiI (ggccgcctcgcc); Site:2: SfiI (ggccattatgcc); 5' an
      3' adaptors were used in cloning as follows: 5' adaptor
      sequence: 5'-CAGGGCATATATGCCC-3' and 3' adaptor sequence
      5'-ATTCTAGAGCCGAGGCGCCGACATG-dt(30)BN-3' (where B = A,
      C, or G and N = A, C, G, or T). Average insert size 1.9
      kb (range 0.5-4.0 kb). 12/15 colonies contained inserts
      by PCR. This library was enriched for full-length clones
      and was constructed by Clontech Laboratories (Palo Alto,
      CA). Note: this is a NIH_MGC Library."
      376 a 187 c 280 g 184 t
BASE COUNT

```

Query Match 3.5%; Score 194.8; DB 12; Length 1027;

[illegible]

RESULT	10
BQ674146	
LOCUS	linear EST 15-JUL-2002
DEFINITION	AGENCOURT_8414352 NIH_MGC_102 Homo sapiens cDNA clone IMAGE:6272500
ACCESSION	BQ674146
VERSION	5', mRNA sequence.
KEYWORDS	BQ674146
SOURCE	BQ674146.1 GI:21784980
ORGANISM	human.
REFERENCE	Homo sapiens
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: cgabbs-r@mail.nih.gov Tissue Procurement: ATCC cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLCM2450 row: p column: 05 High quality sequence stop: 515.

[illegible]

BASE COUNT  
ORIGIN

```
Query Match          3.5%; Score 194.4; DB 14; Length 1016;
Best Local Similarity 85.0%; Pred. No. 1.1e-12;
Matches 216; Conservative 1; Mismatches 37; Indels 0; Gaps 0;
```



[illegible]

Qy	5480	TTGCCCAGGCTGGTCTCAAACTCCTGACCTCAGA	5513
Db	120	TTGCCCAGGCTGGTGTCAAACTCCTGTCTCAGA	87
RESULT 12			
	AQ356384	532 bp	linear
LOCUS	AQ356384		GSS 24-JAN-1999
DEFINITION	CITBI-EI-2529E11.TR CITBI-EI Homo sapiens genomic clone 2529E11,		
	DNA sequence.		

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 532)  
Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and  
Venter,J.C.  
Use of BAC End Sequences from CalTech Libraries for Sequence-Ready  
Map Building  
Unpublished (1997)  
Other\_GSSs: CITBI-El-2529E11.TF  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbeetigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: M13 Reverse

```

seq primer. was reverse
Class: BAC ends.
FEATURES             source
    Location/Qualifiers
        1..532
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_ref="2529E11"
            /clone_lib="CITBI-E1"
            /sex="male"
            /cell_type="sperm"
            /note="Vector: pBelOBAC11; Site_1: EcoRI; Site_2: EcoRI;
            Caltech Human BAC Library D"
BASE COUNT          103 a   134 c   110 g   185 t
ORIGIN

Query Match          3.5%;   Score 194.2;   DB 17;   Length 532;
Best Local Similarity 85.9%;   Pred. No. 1.8e-12;

```

```

site and the (dT)18 tail. The sequence tag for this
library is GTTCTACGAG.
TAG_LiB=UI-H-DP0
TAG_TISSUP=fibrosarcoma
TAG_SBO=CTTCTACGAG"

BASE COUNT      104 a      111 c      82 g      154 t
ORIGIN

Query Match      3.5%; Score 194; DB 14; Length 451;
Best Local Similarity 86.9%; Pred. No. 2.1e-12;
Matches 212; Conservative 1; Mismatches 31; Indels 0; Gaps 0;

Qy 5269 TTTTTTTTTTTTTTTTTTTTTTTTTCAGATGGAGTTTTGGTCTGTGTGCCACTGCTGGAG 5328
Db 170 TTTGATCTTTTTTTTTTTTTTTTTTTTTCAGAGCGAGTTTTGCTCTGTGTGCCACGGCTGGAG 229
Qy 5329 TGGAAATGGCATGAYCATAGCTCACTGCAACTCCAGCTCCTCGGGTTTCAAGCAAGCTGTC 5388
Db 230 TGCATGGCGCCATCTTGGCTCACTGCAACTCCGCTCTCTGGTTCAAGCGATTCTCCT 289
Qy 5389 GCCTCAGCCTCCCGGGTAGATGGGATTACAGCGCCGCCACACCACTCGGTAAATGTTTT 5448
Db 290 GCCTCAGCCTCCCGAGTAGCTGGGGATTACAGGCATGCACCAACCCAGCCAGCTAAATTTT 349
Qy 5449 GTATTTTTAGTAGAGATGGGGTTTTCACCATGTTTGGCCAGCGTGGTCTCAAACTCCCTGACC 5508
Db 350 GTATTTTTAGTAGAGATGGGGTTTTCACCACATTGGCCAGCGTGGTCTCAAACTCCCTGACC 409
Qy 5509 TCAG 5512
Db 410 TCAG 413

```

Db	290	GCCTCAGCCTCCCGAGTAGCTGGGATTACAGGCATGCACACACGCCAGCTAATTTTT	349
Qy	5449	GTATTTTTAGTAGATGGGGTTTACCACATGTTGGCCAGGCTGGTCTCAAACTCCTGACC	5508
Db	350	GTATTTTTAGTAGATGGGGTTTACCACATGCGCAGGCTGGTCTCAAACTCCTGACC	409
Qy	5509	TCAG 5512	
Db	410	TCAG 413	

```

QY      5509 TCAG 5512
       ||||
Db      410 TCAG 413

RESULT 14
BF575954/c
LOCUS   BF575954                885 bp      mRNA      linear      EST 12-DEC-2000
DEFINITION  G02132930F1 NIH_MGC_81 Homo sapiens cDNA clone IMAGE:428254 5',
              mRNA sequence.
ACCESSION  BF575954
VERSION    BF575954
KEYWORDS   GI:11649666
           EST.
SOURCE     human.

```

```

CDNA Library Preparation: CLONETECH Laboratories, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLC1127 row: k column: 07
High quality sequence stop: 302.
Location/Qualifiers
1. 885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4288254"
/clone_lib="NIH_MGC_81"
/lab_host="DH10B (T1 phage-resistant)"
/note="Organ: muscle (skeletal); Vector: pDNR-LIB
(Clontech); Site_1: SfiI (ggccgctggcc); Site_2: SfiI
(ggccattatggcc); 5' and 3' adaptors were used in cloning
as follows: 5' adaptor sequence: 5'-CACGCCATTATGGCC-3'
and 3' adaptor sequence:
5'-ATTCTAGAGCCGAGGCGGCCACATG-dT(30)BN-3' (where B = A,

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 19, 2003, 00:01:09 ; Search time 231.577 Seconds  
(without alignments)  
7302.178 Million cell updates/sec

Title: US-09-966-880A-9  
Perfect score: 5514  
Sequence: 1 acagacgaatacatagtgtcca.....tcaaactctgacctcgag 5514

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PTCUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	218.6	4.0	4421	2	US-08-257-963B-9
C 2	218.6	4.0	4421	4	US-08-367-841A-9
C 3	218.6	4.0	4421	4	US-08-520-373D-6
C 4	218.6	4.0	4421	5	PCT-US95-07201-9
C 5	199.2	3.6	246240	2	US-08-724-394A-20
C 6	199.2	3.6	246240	2	US-08-724-394A-21
C 7	199.2	3.6	246240	2	US-08-724-394A-22
C 8	195	3.5	17327	1	US-07-906-871-15
C 9	194.8	3.5	72928	3	US-09-009-913-1
C 10	193.8	3.5	43950	4	US-09-735-934A-3
C 11	193.8	3.5	98844	4	US-09-791-211-10
C 12	191	3.5	29629	4	US-09-729-995-3
C 13	190.8	3.5	98844	4	US-09-791-211-10
C 14	190.2	3.4	111282	4	US-09-754-250-3
C 15	188.2	3.4	168575	4	US-09-426-290-1
C 16	188	3.4	2561	4	US-09-270-542-101
C 17	188	3.4	2561	4	US-09-270-542-119
C 18	187.8	3.4	17327	1	US-07-906-871-15
C 19	187.4	3.4	9365	4	US-08-608-285A-8
C 20	187.4	3.4	9365	4	US-09-350-836B-8
C 21	187.4	3.4	9365	4	US-09-370-265-8
C 22	187.4	3.4	14747	4	US-09-608-285A-42
C 23	187.4	3.4	15977	4	US-09-608-285A-59
C 24	187	3.4	15297	4	US-09-817-180-3
C 25	186.6	3.4	112132	4	US-09-741-150-3
C 26	186	3.4	43950	4	US-09-735-934A-3
C 27	184.8	3.4	80246	4	US-09-078-294-4

C 28	184.8	3.4	99500	4	US-09-798-096-10	Sequence 10, Appl
C 29	184.4	3.3	81001	4	US-09-750-580-1	Sequence 1, Appli
C 30	184.4	3.3	81001	4	US-09-750-580-1	Sequence 1, Appli
C 31	184	3.3	72604	4	US-09-268-992-7	Sequence 7, Appli
C 32	184	3.3	72604	4	US-09-657-474-7	Sequence 7, Appli
C 33	183.6	3.3	162450	4	US-09-345-882-1	Sequence 1, Appli
C 34	183.4	3.3	2561	4	US-09-270-542-101	Sequence 101, App
C 35	183.4	3.3	2561	4	US-09-270-542-119	Sequence 119, App
C 36	183.4	3.3	31571	1	US-08-323-443B-1	Sequence 1, Appli
C 37	183.4	3.3	53526	3	US-08-658-136-2	Sequence 2, Appli
C 38	183.4	3.3	53577	3	US-08-658-136-1	Sequence 1, Appli
C 39	183	3.3	35060	3	US-08-814-095-7	Sequence 7, Appli
C 40	182.6	3.3	36159	4	US-09-749-588-3	Sequence 3, Appli
C 41	182.6	3.3	99500	4	US-09-798-096-10	Sequence 10, Appl
C 42	182.4	3.3	14581	4	US-08-520-373D-4	Sequence 4, Appli
C 43	182.4	3.3	14581	4	US-08-520-373D-4	Sequence 4, Appli
C 44	182.4	3.3	22481	4	US-08-367-841A-43	Sequence 43, Appl
C 45	182.4	3.3	22481	4	US-08-367-841A-43	Sequence 43, Appl

ALIGNMENTS

RESULT 1  
US-08-257-963B-9/c  
; Sequence 9, Application US/08257963B  
; Patent No. 5840686  
; GENERAL INFORMATION:  
; APPLICANT: Chader, Gerald J.; Becerra, S.  
; APPLICANT: Patricia; Schwartz, Joan P.;  
; APPLICANT: Taniwaki, Takayuki  
; TITLE OF INVENTION: PIGMENT EPITHELIUM  
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION OF ITS NOVEL  
; TITLE OF INVENTION: BIOLOGICAL ACTIVITY AND SEQUENCES ENCODING  
; TITLE OF INVENTION: AND EXPRESSING THE PROTEIN  
; NUMBER OF SEQUENCES: 42  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Morgan & Finnegan  
; STREET: 345 Park Avenue  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy Disk  
; COMPUTER: IBM PC Compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WORDPERFECT 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/257,963B  
; FILING DATE:  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 07/952,796  
; FILING DATE: 24-SEPT-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: DOROTHY R. AUTH  
; REGISTRATION NUMBER: 36434  
; REFERENCE/DOCKET NUMBER: 20264126US1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 758-4800  
; TELEFAX: (212) 751-6849  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4421 Base Pairs  
; TYPE: Nucleic Acid  
; STRANDEDNESS: Double  
; TOPOLOGY: Unknown  
; MOLECULE TYPE: Genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: JT101



```

RESULT 2
US-08-367-841A-9/c
; Sequence 9, Application US/08367841A
; Patent No. 6319687
; GENERAL INFORMATION:
; APPLICANT: Chader, Gerald J.; Rodriguez,
; APPLICANT: Ignacio R.; Mazuruk, Krzysztof;
; APPLICANT: Tombran-Tink, Joyce
; TITLE OF INVENTION: PIGMENT EPITHELIUM
; TITLE OF INVENTION: DERIVED FACTOR: CHARACTERIZATION GENOMIC
; TITLE OF INVENTION: ORGANIZATION AND SEQUENCE OF THE PDF GENE
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Morgan & Finnegan
; STREET: 345 Park Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/367,841A
; FILING DATE: 30-DEC-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/257,963
; FILING DATE: 07-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/952,796
; FILING DATE: 24-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36434
; REFERENCE/DOCKET NUMBER: 20264126US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4421 Base Pairs
; TYPE: Nucleic Acid
; STRANDEDNESS: Double
; TOPOLOGY: Unknown
; MOLECULE TYPE: Genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Human
; FEATURE:
; NAME/KEY: JT1
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: 7.1 kb Bam HI
; OTHER INFORMATION: fragment Derived from human placental
; OTHER INFORMATION: genomic DNA; Also referred to as JT101
US-08-367-841A-9

Query Match      4.0%; Score 218.6; DB 4; Length 4421;
Best Local Similarity 47.7%; Pred. No. 6,8e-37;
Matches 952; Conservative 1; Mismatches 995; Indels 47; Gaps 9;

QY      3538 ATTTAAAGTGAACAGACAGCCAGGTGCTGGCCTTGAGCCCTGAGCTTCAGACCAGCGCTGAGCAACATGG 3597
          ||||| |   ||||| |   ||||| |   ||||| |   ||||| |   ||||| |   ||||| |   ||||| |
Db      4107 AGTTAAAGAGGTTTTGCTGGCCAGGTATGCTGGCTCACGCCTGTAAATCCAGCACTCTGG 3988

QY      3598 GAGGCTGAGGTGGGTGGATCGCTTGAGCCCTGAGCCCTGAGCTTCAGACCAGCGCTGAGCAACATGG 3657
          ||||| |   ||||| |   ||||| |   ||||| |   ||||| |   ||||| |   ||||| |   ||||| |
Db      4047 GAGGCTAAGGTGGCGGATACACCTGATGTGTCAGAGTTCAAGACCGAGCTGCCAACATGG 3988

```

Db 2930 TCCAGTCCCGAGAAAAGACTTAAACACTAGTATCTGGAGATCCCAAGTCTCTTTAAAGGA 2871  
QY 4778 CATCACTTGTGCTTTTCATCAACACGCTGTACAGCTTGTGTGCTGTCTGCAGCTGCA 4837  
Db 2870 TTTAGCTGATATGACGAGCATCGATGTACCTGGGGTTTGGCTGGATCCACCTGG 2811  
QY 4838 ATGGGACCTCTGATTTCTTTAAGAAACTTGGGTACAGAGTATTTCCACAAATGCTTA 4897  
Db 2810 CTCACTGATGATGCTGTCTTCCCTGGAACATTCCTGTGTTTGTAGTAGTATGAGATGT 2751  
QY 4898 TTCAAATTTAGTCTTATGATATGCAAGACACTGTGCTAGGAGCCAGAAACAAAGAGAG 4957  
Db 2750 GTGCACCACTCCAGTGGGAACCTGAAGCTCTTAGTAGTTACATTTGTAATTTAAAGAAAT 2691  
QY 4958 GAGAAATCACTATGTTGGGAACAACATACAGATATTTAGATCAATTTTGACTAGTT 5017  
Db 2690 ATAAACAGAAATCAGCAAGCTGCTTACCTCCAGAGACACACCGTTCACCTTACAGTA 2631  
QY 5018 AAAAAAGCAGAGTACAAATACACATGCAATCAATCAATCAATCAATCAATCAATCAAT 5077  
Db 2630 TCTAGCTTCTCGACTCCCTTGTAAATCGTGGAGTATGAACCTGTGACCAAGAGTAA 2571  
QY 5078 TGTGCTGTAGAAAGACTAGAGAAATAACACAGAACTTAAACAGTCAATTTGCTATTAGA 5137  
Db 2570 ATATGAGCGGTTTATTGAGCAGTTTCTCCAGGGATGACGGCTGTCACTGTGTCTTCT 2511  
QY 5138 CACTAAGCTAATTTATTTATTTAGACACTATGATTTGAGATTTAAATAATCTTTAAT 5197  
Db 2510 GCCAGTTTCCAGTCTTACTCTCTGCCCTTCTATCTGTACTCTTGGCAGAGTCAGTAAC 2451  
QY 5198 ATTTTAAATTTAGAGCTCTCTATTTTCCATAGTATTTCAAGTTTGCAATGATCAAGT 5257  
Db 2450 CAACCACTAATAAAATC---AAAGTTGTAGTAAAGACTTTATAACACATATGCCCTTC 2394  
QY 5258 ATACTCTTCTTT 5317  
Db 2393 TGAGTTGTGGTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 2334  
QY 5318 CCATGCTGGAGTGGATGTCATGAYCATAGCTCACTGCAACCTCCACCTCCCTGGGTTCAA 5377  
Db 2333 CCAGGCTGAGCGCAATGCGCGATCTTGGCTTACGCCAATCCGCCCTCCGGGTTCAA 2274  
QY 5378 GCAAGAGCTGCGCTCAGCTCCCGGGTAGATGGGATACAGCGCCCAACCAACCACTC 5437  
Db 2273 GTGATTTCTCTGCTCAGCTCCCGAGTAGCTGGGATTAACAGCGCCGCCCAACCACT 2214  
QY 5438 GGCTAATGTTCTATTTTATGATAGATGGGTTTCCACCATGTTGCCAGGCTGCTCA 5497  
Db 2213 GGCTAAT-PTTGATTTTATGATAGACGGGATTTTCACTATGTTGGTCAGGTTGGTATCA 2155  
QY 5498 AACTCTGACCTCAG 5512  
Db 2154 AACTCCCGCCCTCAG 2140

RESULT 3  
US-08-520-373D-6/c  
; Sequence 6, Application US/08520373D  
; Patent No. 6451763  
; GENERAL INFORMATION:  
; APPLICANT: Tombran-Pink, Joyce  
; APPLICANT: Steele, Fintan R  
; APPLICANT: Chader, Gerald J  
; APPLICANT: Becerra, Sofia P  
; APPLICANT: Johnson, Lincoln V  
; APPLICANT: Rodriguez, Ignacio R  
; TITLE OF INVENTION: RETINAL PIGMENTED EPITHELIUM DERIVED NEUROTROPIC FACTOR  
; FILE REFERENCE: 2026-4203US1  
; CURRENT APPLICATION NUMBER: US/08/520, 373D  
; CURRENT FILING DATE: 1995-08-29  
; PRIOR APPLICATION NUMBER: 08/377,710  
; PRIOR FILING DATE: 1995-01-25

; PRIOR APPLICATION NUMBER: 08/279, 979  
; PRIOR FILING DATE: 1994-07-25  
; PRIOR APPLICATION NUMBER: 07/894, 215  
; PRIOR FILING DATE: 1992-06-04  
; PRIOR APPLICATION NUMBER: 07/952, 796  
; PRIOR FILING DATE: 1992-09-24  
; NUMBER OF SEQ ID NOS: 34  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 6  
; LENGTH: 4421  
; TYPE: DNA  
; ORGANISM: HUMAN  
; FEATURE:  
; OTHER INFORMATION: CDS 66-322  
US-08-520-373D-6  
  
Query Match 4.0%; Score 218.6; DB 4; Length 4421;  
Best Local Similarity 47.7%; Pred. No. 6.8e-37;  
Matches 952; Conservative 1; Mismatches 995; Indels 47; Gaps 9;  
  
QY 3538 ATTTAAAGTGAAACACAGACAGCCAGGTGTGTGGCTCACGCCCTGTATATCCAGCACTCTGG 3597  
Db 4107 AGTTAAAGAGGTTTGTCTGCCAGGTATGTGCTCACGCCCTGTATATCCAGCACTCTGG 4048  
QY 3598 GAGGCTGAGGTGGGTGGATCGCTTGAGCCCTGGAGTTCAAGACCAACCCCTGAGCAACATGG 3657  
Db 4047 GAGGCTAAGGTGGGGGATCACTGTATGTCAAGAGTTCAAGACCAACCCCTGAGCAACATGG 3988  
QY 3658 CAAAACCCCTGTTCT-----ATACAAAAAATTAGCCGGCATGTGCGCATGTGCTCT 3710  
Db 3987 TGAACCCCATCTTTACTTAAATAACAAAACTAGCCAGCGGTGTGTTGGGCACCTGT 3928  
QY 3711 GGTCCAGCTACTTAGG-GGCTGAGCAGGAGAACTCTTTGGAGCCCGAGAGGTCAAGGCT 3769  
Db 3927 AATCCAGCTACTTGGAGGCTGAGCAGGAGAACTCTTTGAACCCAGCAGCGGAGGTT 3868  
QY 3770 GCAGTGACAGTGTGTGGCCCACTGCATCCAGCCCTGGGTGACAGACCAAGCACTTGGCT 3829  
Db 3867 GCAGTGACCGAGATGCTCTCCACTGCCACTCCAGCCCGAGGTGACAGTGTGAGACTCCATCT 3808  
QY 3830 C-AAAAAATAAGAGAAAAAATTTAAAAATAATGGAACAACTACAAACAGCTGTTGTCC 3888  
Db 3807 CAAAAAATAAGAGCGCGGTGCGGTGCGTCAAGCTGTATATCCAGCACTTTGGG 3748  
QY 3889 TAGTAGAGCTACTTATAGTGTATTTTGGTATTTTAACTTTTAAAGTCAGGCTGT 3948  
Db 3747 AGCCGAGCGCAGCATCACGAGGTCCAGAGATCCAGACCATCTCTGGCTAACACGCTCA 3688  
QY 3949 CACCTGCATCATTTATTAATAATATCAATTTCTCAATGTATATCCACACAAAGCACTGGTAC 4008  
Db 3687 AACCCCGTCTCTACTAAAAA-----TACAAAAAATTTAGTCG 3651  
QY 4009 GTGAATGTTTCATAGTACCTTTATTCACAAAAACCCCAAGTAGAGACTATCCAAATATCCA 4068  
Db 3650 GCGGTGGTGGCGAGCGCTGTATTCACACTACTCGGAGGTTGAGGCAAGGACAATGGG 3591  
QY 4069 TCAACAAGTGAACAAATAAACAATAATGTCTATATCCATGCAATGGAATACCAACCTGCA 4128  
Db 3590 TGAACCCGGGAGGAGAGGCTTGCAGTGGCCGAGATCGGCCACTGCATCTCCAGCCTGGG 3531  
QY 4129 GTACAAAGGAAGAAGCTACTTGGGGATGAATCCCAAGTCATGACGCTAAATGAAGAAGT 4188  
Db 3530 TGACAGAGCCAGACTCCATCTCAAGAAAAAATAAATAAATAAAGTCTGCTGCCAG 3471  
QY 4189 CAGACATGAAGGAGGAGATAATGTATGCCATACGAAATTTCTAGAAATCAAAAGTAAC--- 4245  
Db 3470 ATGTGGTTCTGTGGTGGGTGCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 3411  
QY 4246 --TTATAGTTACAGAAAGCAAAATCAGGCGAGCATAGAGGCTCACACCTGTATATCCAGC 4303  
Db 3410 GTGCGGATCTGTGTAGGTTCTGCAGGCCAGGTGATAGAGGCTCACACCTGTATATCCAGT 3351  
QY 4304 ACTTTGAGAGGCC-ACGTGGGAAGATTGCTAGAACTCAGAGTTCAAGACCAAGCCCTGGGC 4362





Best Local Similarity 47.7%; Pred. No. 6.8e-37; Matches 952; Conservative 1; Mismatches 995; Indels 47; Gaps 9;			
QY	3538	ATTAAAGTGAACACAGCAGCGAGGTGGTGGCTCAGCGCTGTAAATCCCGACCTCTGG	3597
Db	4107	AGTTAAAGAGGTTTGTCTGGCCAGGTATGGTGGCTCAGCGCTGTAAATCCCGACCTTGG	4048
QY	3598	GAGGTGAGGTGGGTGGATTCGTTGAGCCCTGGAGTTCAAGACCCAGCCTGAGCAACATGG	3657
Db	4047	GAGGTAAAGTGGGGGATCCCTGATGTCAGGAGTTCAAGACCCAGCCTGGCCACATGG	3988
QY	3658	CAAAACCTGTTCT- - - - - ARAACAAAATAGCCGGCATGGTGCATGTCCTGT	3710
Db	3987	TGAACCCCACTCTTACTAAATAACAAAACCTAGCCAGCGTGGTGGGACCTGT	3928
QY	3711	GGTCCAGCTACTAGG- -GGTGGAGCAGGAGATCTTTGGAGCCAGGAGGTCAAGCT	3769
Db	3927	AATCCAGCTACTTGGAGGCTGAGCAGGAGATCTCTGNACCCAGGAGCGAGGT	3868
QY	3770	GCAGTGCAGTGTGGCCGCACTGCATCCAGCCTGGGTGACAGGACAGACCTTGGCT	3829
Db	3867	GCAGTGCAGGAGATCGCTCCACTGCATCCAGCCAGGTGACAGTGTGAGACTCCATCT	3808
QY	3830	C-AAAAAATAGAGAAAATTAATAATGAACAACTACAAAGAGCTGTGTCC	3888
Db	3807	CAAAAAAATAAGAGCCGGTGGCTCAAGCCTGTAATCCAGCAGCTTTGGG	3748
QY	3889	TAGATGAGCTACTTGTAGTGTGATATTTTGGTATTTAACTTTTAAAGTCAGGTCGT	3948
Db	3747	AGGCGAGCGGAGATCAGGAGTCAGGAGTCCAGACCATCTTGGCTAACAGGTGA	3688
QY	3949	CACCTGCATCATATTAAATATCAATTCATATATCCACAAAAGACTGGTAC	4008
Db	3687	AACCCGCTCTACTAAAAA- - - - - TACAAAAAATTAGTCG	3651
QY	4009	GTGAATGTCATAGTACCTTTATTCACAAAACCCAAAGTAGAGACTATCCAAATATCA	4068
Db	3650	GGGTGGTGGCGGAGCCTGTATTCACAGCTACTCGGAGGTGAGGAGGAGAAATGGG	3591
QY	4069	TCAACAAGTGAACAAATAAACAAAATGTCTATATCCATGCAATGGAATACCACTGCA	4128
Db	3590	TGAACCCGGAGGAGAGCTTGCAGTGAGCCGAGATCGCCACTGCACTCCAGCCTGG	3531
QY	4129	GTACAAAGAGAGAGCTACTTGGGGATGAATCCCAAGTCAATGAGCTAAATGAAGAGT	4188
Db	3530	TGACAGACCCAGACTCCATCTCAAAAGAAAAAATAATACTAAGAGGTCTCTGTCGCA	3471
QY	4189	CAGACATGAAGGAGGAGATAATGATGCCATACGAAATTCAGAAAAATGAAGTAAC- - -	4245
Db	3470	ATGTGGTCTGTGGTGGGTGGTCTGTGGTGGGTGGTGGTCTGCGAGCCAG	3411
QY	4246	- -TTATAGTTACAGAAACAAATACAGGCGAGGCATAGAGGCTCACACCTGTAAATCCAGC	4303
Db	3410	GTGCGGATCTGTGTAGTCTTGCAAGGCCAGTGTAGAGGCTCACACCTGTAAATCCAGT	3351
QY	4304	ACTTTGAGAGGCC- -ACGTGGGAAGATTCGTAGAACTAGAGGTTCAGAACCCAGCCTGGC	4362
Db	3350	ACTTTGAGAGCGTGAAGTGGAGGATCACTTGAAGCCAGGAGTTCAAAACCCAGCCTGGC	3291
QY	4363	AACAGTGAACCTCCATCTCCACAAAATGGGAAAAAAGAAAGCAAAATCAGTGGTTG	4422
Db	3290	AATATGTGAGACCGCTATCACTACAAAAAGTTTGTGTACCCGAGCATGGTGGCAC	3231
QY	4423	TCCTGTGGGAGGAGGAGCTCAAGAGGGAAGAGCTCTGTGGGTGAGGGTGGT	4482
Db	3230	ATGCTGTAGTCCCACTACTCGAGAACTGAGATAGGCTCACTGACCTGGTAGGCTG	3171
QY	4483	ATTC- - - - - AGTTCTGTATCCTGACTGTGTGAGCAGTTTGGGGTGTATACATCAAAA	4537
Db	3170	AGGTGCACTGAGCAAGATCCGACTACTGCACCTCCAGCCTGGGTGACAGAGTAAGACCC	3111
QY	4538	TATCGTAGAATTATGCATCTTAATGGGTGAGTTTACTGTATGTAATATACCTCAA	4597

RESULT 5  
US-08-724-394A-20/C  
; Sequence 20, Application US/08724394A  
; Patent No. 5872237  
; GENERAL INFORMATION:  
; APPLICANT: Feder, John N.

Db	3110	TGCTCTCAACAAAAAATAACAGGTTCTGTTGCTGCTGAGCTTTGCTCAGGAAGATAAGAC	3051
QY	4598	TGTAAAGAAAAATAATGTGTAAAGAAAGTTTCAATTTCTTTGCCAGCAACAGCTTATTCAA	4657
Db	3050	TTTCCCAACAGTCCGAAGCCTAGGATATGAGGGCTGAGGAGGACTCAAAACCGTTGCAGTC	2991
QY	4658	ATTCCTGAGCCCTTTTACTTTCGCAAAATCTCTGCACCTTCTGCCCGCTACCATTTAGTGACA	4717
Db	2990	AGAGAGACACCTATTTTGGTGAATATTTTGAATAGTTAAAGTGGCCTCCAGTACAAC	2931
QY	4718	GCACCTAGCTCCACAAAATGGGATAAATGCAATTTCTGAAAAAGACTAGGGAACAAATCCAGG	4777
Db	2930	TCCCAGTTCCCAGAAAAGACTTAAAACTAGTATCTGGAGATTCCAAGTCTTTAAAGGA	2871
QY	4778	CATCACTTGTGCTTTCATATCAACCCAGCTGTACAGCTTGTGTGCTGTCTGCAGCTGCA	4837
Db	2870	TTTTAGCTGATATGAGGCATCTGCATGTTAGCCTGGGTTTGTGGCTGGATTCACCTCG	2811
QY	4838	ATGGGAGCTCTTGATTTCTTTAAGGAAACTTGGGTTTACCAGAGTATTTCCACAAATGCTA	4897
Db	2810	CTCAGTGATGATGATGCTGTTTCCCTGGAACATTCCTGTGGTTTATAGTAAATGAGATGT	2751
QY	4898	TTCAAAATTAGTGCTTATGATATGCAAGACACTGTGCTAGGAGCCAGAAAACAAAGGAG	4957
Db	2750	GTGCACCACTCCAGTGGGAAGCTTGTAGTAGTTACATTTGTAATTTAAAAAAGAATTT	2691
QY	4958	GAGAAATCAGTCATTTATGTGGGAACAACATAGCAAGATATTTTAGATCATTTTGTAGCTGT	5017
Db	2690	ATAAAAAAGAAATCAGCAAGCCTGCTTACCCTCCAGAGACACACCTTCCACTTACAGTA	2631
QY	5018	AAAAAGCAGCAGATACAAAATCACATGCAATGCAATCAGTATTAATCCAAATCATGTAAATA	5077
Db	2630	TCTAGCTTCTCGACTCCTTTGTTAAAAATCGTGGAGTATGAACCTGTGACAGAGAGTAA	2571
QY	5078	TGTGCTCTGAGAAAGACTAGAGGAATAAACCAAGAACTTTAAACAGTCACTTCTCATTAGA	5137
Db	2570	ATATGGAGCGGTTTATTTAGAGCAGTTTCCCTCAGGAGTACGCGCTGTCACTGTGTTCCT	2511
QY	5138	CACTAAGTCTAATTTATTTATTTAGACACTATGATATTTGAGATTTTAAAAATCTTTTAA	5197
Db	2510	GCCAGTTTTCCAGTCTTACTCTCGCCTTCATTTCTGTACTCTTGGCAGAGCTCAGTAAC	2451
QY	5198	ATTTTAAATTTAGAGCTCTCTCTATTTTCCATAGTATTTCAAGTTTGTGACATGATCAAGT	5257
Db	2450	CAACCACTAATAAAAAATC- - - AAAGTTGTAGTAAAGACTTTTAAACACATATGCCCTTC	2394
QY	5258	ATTACTCTTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTAGATGGAGTTTGGTCTTGTTC	5317
Db	2393	TGAGTTGTGGTTTTGTTTTGTTTTGTTTTGTTTTGTTTTGAGATGGAGTTTCACTCTTGTTC	2334
QY	5318	CCATGCTGGAGTGAATGGCATGAYCATAGCTCACTGCAACCTCCACCTCTCGGGTTCAA	5377
Db	2333	CCAGGCTGGAGGCAATGGCGGATCTTGGCTTACGGCAACCTCCGCCCTCCGGGTTCAA	2274
QY	5378	GCAAGCTGTGCGCTCAGCCTCCCGGGTAGATGGGATTTACAGGCGGCCACACCACTC	5437
Db	2273	GTGATTTCTCTGCTTCCAGCTCCCGAGTGTGGGATTTACAGGCGGCCGCCACCACTC	2214
QY	5438	GGCTAATGTTGTTATTTTGTAGAGATGGGTTTCCACCATGTTGCCAGGCTGTGCTCA	5497
Db	2213	GGCTAAT- -TTGTATTTTGTAGTAGAGCGGATTTTCACTATGTTGTCAGGTTGGTATCA	2155
QY	5498	AACTCCTGACCTCAG	5512
Db	2154	AACTCCCGCCTCAG	2140